DISSERTATIONES BIOCENTRI VIIKKI UNIVERSITATIS HELSINGIENSIS

9/2013

TIINA RASILA

Functional Mapping of Mu Transposition Machinery: MuA Protein Modification and Engineering for Hyperactivity

INSTITUTE OF BIOTECHNOLOGY AND
DIVISION OF BIOCHEMISTRY AND BIOTECHNOLOGY
DEPARTMENT OF BIOSCIENCES
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES
UNIVERSITY OF HELSINKI

Functional Mapping of Mu Transposition Machinery: MuA Protein Modification and Engineering for Hyperactivity

Tiina Rasila

Institute of Biotechnology
and
Division of Biochemistry and Biotechnology,
Department of Biosciences,
Faculty of Biological and Environmental Sciences,
University of Helsinki

Academic dissertation

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in auditorium 2 in Viikki Infocenter, Viikinkaari 11, Helsinki, on May 17th 2013 at 12 o'clock noon.

Helsinki 2013

Supervisor Professor Harri Savilahti
Department of Biology
Faculty of Mathematics and Natural Sciences
University of Turku
Turku, Finland

Reviewers Docent Suvi Taira
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

Docent Tuomo Glumoff
Department of Biochemistry
Faculty of Science
University of Oulu
Oulu, Finland

Opponent Professor Jari Valkonen
Department of Agricultural Sciences
Faculty of Agriculture and Forestry
University of Helsinki
Helsinki, Finland

Custos Professor Jukka Finne
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

ISBN 978-952-10-8718-9 (paperback)
ISBN 978-952-10-8719-6 (PDF)
ISSN 1799-7372

Press: Unigrafia, Helsinki 2013

LEARN FROM YESTERDAY, LIVE FOR TODAY, HOPE FOR TOMORROW.
THE IMPORTANT THING IS TO NOT STOP QUESTIONING.
–Albert Einstein

Contents

Original publications

This thesis is based on the following articles, which are referred to in the text by their Roman numerals.

I Tiina Rasila, Maria Pajunen, and Harri Savilahti, (2009) Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Anal. Biochem.*, 388, 71-80. Reprinted with the permission from the publisher.

II Maria Pajunen, Tiina Rasila, Lotta Happonen, Arja Lamberg, Saija Haapa-Paananen, Saija Kiljunen, and Harri Savilahti, (2010) Universal platform for quantitative analysis of DNA transposition. *Mobile DNA,* 1, 24.

III Tiina Rasila, Mauno Vihinen, Lars Paulin, Saija Haapa-Paananen, and Harri Savilahti, (2012) Flexibility in MuA transposase family protein structures: Functional mapping with scanning mutagenesis and sequence alignment of protein homologues *PlosOne,* 7, 5:e37922.

IV Tiina Rasila, Mauno Vihinen, Lars Paulin, Phoebe Rice, and Harri Savilahti, Hyperactivity-probing of Mu transposition recombination machinery, manuscript

Unpublished data will also be presented.

Abbreviations

8-oxo-dGTP	8-oxo-2'-deoxyguanosine-5'-triphosphate
aa	amino acid
ATP	adenosine triphosphate
bp	base pair(s)
CDC	cleaved donor complex
cDNA	complementary deoxyribonucleic acid
C-terminal	carboxy-terminal
DDE	aspartic acid, aspartic acid, glutamic acid (DDE motif)
DNA	deoxyribonucleic acid
dPTP	2'-deoxy-P-nucleoside-5'-triphosphate
ep	error prone
FPLC	fast protein liquid chromatography
hESC	human embryonic stem cell
HIV	human immunodefiency virus
HPLC	high-pressure liquid chromatography
HTH	helix-turn-helix
HU	*E. coli* integration host factor
IAS	internal activating sequence (in Mu genome)
IHF	*E. coli* integration host factor
iPSC	induced pluripotent stem cell
IPTG	isopropyl-β-D-1-thiogalactopyranoside
IS	insertion sequence
LE	left end
LER	left end – enhancer – right end; synaptic complex in Mu transposition
LINE	long interspersed repeated element
LTR	long terminal repeat
MITE	miniature inverted-repeat transposable element
MuA	bacteriophage Mu transposase protein A
MuB	bacteriophage Mu transposition protein B
NH_2OH-HCl	hydroxylamine hydrochloride
N-terminal	amino-terminal
OH	hydroxide
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PFV	prototype foamy virus
RE	right end
RNA	ribonucleic acid
RNase H	ribonuclease H
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
SINE	short interspersed repeated element

SNP	single-nucleotide polymorphism
SSC	stable synaptic complex
TE	transposable element
TIM	transposon integration mediated mutagenesis
TIR	terminal inverted repeat
TriNex	trinucleotide change
V(D)J	variable (diversity) joining; (antigen receptor gene segments)
wt	wild type
X-gal	5-bromo-4-chloro-3-inodyl-β-D-galactopyranoside
YREK	tyrosine, arginine, glutamic acid, lysine (YREK motif)

Abstract

Transposable DNA elements constitute a class of discrete genome segments, which use fundamentally similar reactions for their movement within and between genomes. The similarity of the reaction mechanisms is reflected by their enzymes, the transposases. Many DNA transposases share a conserved RNase H-like fold that contains a catalytic DDE motif. The bacteriophage Mu encodes MuA transposase, which is a well characterized member of the DDE transposases. Mu transposition proceeds within a nucleoprotein complex known as a transpososome, the core of which contains four MuA molecules and two synapsed transposon ends. The transpososome machinery *in vivo* involves auxiliary factors, but *in vitro* this minimal configuration forms the basis of Mu transposon tools that are widely used in genetics/genomics applications. In spite of a detailed knowledge of the chemical reactions of Mu transposition and transposition mechanisms in general, structural information on transpososomes has proven elusive for most of the DNA elements. Structural studies of transposases and functional data provide an unparalleled insight into the mechanistic details of action. These studies are a prerequisite to the understanding of the important biological processes in which they are involved. They are also useful for the evaluation of exploiting DNA transposon as tools.

The present study expands the knowledge of the functional determinants of Mu transposition machinery. Important structure-function relationships of MuA were revealed. Initially, properties for successful random MuA mutant library generation were dissected by comprehensive evaluation of different random DNA mutagenesis methods. Analyses provided useful operational guidelines for generating diverse mutant libraries.

Next, a universal assay that quantitatively analyzes DNA transposition events was established for mutant screening purposes. The assay is based on the inducible expression of transposase, which catalyzes the transposition of a *lacZ* marker gene-containing transposon. Individual transposition events can be scored as blue micro colonies or papillae growing within an otherwise whitish colony. The assay was validated by Mu and IS*903* transposition systems.

The papillation assay was used to analyze MuA mutant libraries generated by two different approaches. First, a pentapetide insertion mutagenesis strategy was used to map MuA protein structure for those regions that withstood insertion without reduction of function. The comprehensive structure-function analysis was complemented with data from the sequence alignment of protein homologues and they yielded a wealth of information about the activity of the transposition machinery. In addition, potential sites for further protein modifications were identified.

Second, different random MuA mutant libraries were screened for increased transposition frequency. MuA mutant variants with altered properties could be generated. Many single-amino acid substitution MuA mutants were tested by two different assays relevant for MuA-based applications. Several MuA mutants were discovered to be valuable for Mu *in vitro* transposition applications. In addition, the mapping of hyperactivity-inducing substitutions to the recently resolved Mu transpososome structure, allowed the factors behind the activity enhancement to be characterized.

The results of this study provide a comprehensive structure-function map of MuA transposase. Mapping of insertion tolerant versus insertion intolerant sites and

hyperactivity-inducing mutations of MuA structure in the context of the entire Mu transpososome, provides fundamental insights into the function of the transpososome. The functional information explains and validates different protein-protein and protein-DNA contacts seen in the crystal structures. They may also provide suggestions for protein conformational changes that accompany the transposition process. Moreover, the activity enhancing substitutions discovered in RNase H-like fold may be generalized to generate hyperactive variants or used for mechanistic analyses of other DDE transposases or used for the analyses of their close relatives such as HIV integrase and the RAG recombinase of the V(D)J recombination.

1 INTRODUCTION

1.1 Transposable elements

Transposable elements (TEs) are discrete segments of DNA that can either move or be copied from one genomic site to another within and between genomes. They use special type of recombination called transpositional recombination or transposition for their movement that is independent of target DNA homology (reviewed by Craig et al., 2002). The transposable elements were discovered by Barbara McClintock, who demonstrated their role in changing the structure of the maize chromosome in the 1940's (reviewed by McClintock 1987). The new concept of genomes as dynamic and fluid entities were received with skepticism and these "jumping genes" were long ignored and called "selfish" or "junk" DNA (reviewed by Plasterk, 1995).

Today, the diverse transposable elements have been identified in all kingdoms of life (reviewed by Campbell, 2002; Feschotte et al., 2002; Filee et al., 2007; Gregory, 2005; Siguier et al., 2006a) and in the genomes of virtually every organism sequenced. TEs are widely accepted important components of genomes and significant contributors of genome evolution (reviewed by Goodier and Kazazian, 2008; Kidwell and Lisch, 2001; Lisch, 2012; Pritham, 2009). They are present in different diversity and copy numbers ranging from just a few elements to tens or even hundreds of thousands per genome (reviewed by Biemont and Vieira, 2005). The number of identified TEs is continuously increasing due to ongoing genome sequencing projects. For the simplest prokaryotic TEs alone, more than 3000 annotated examples are presently known (Siguier et al., 2006b; see IS finder; www-is.biotoul.fr/is.html). TEs are abundant within the eukaryotic genomes. For example, about half of the genome in humans (Lander et al., 2001) is composed of TEs. In contrast to the high proportions found in large genomes, small genomes tend to have a low proportion of TEs (reviewed by Campbell, 2002; Gregory, 2005).

The parasitic lifestyle of TEs has required certain constrains for long-term evolutionary survival (reviewed by Brookfield, 1995; Brookfield, 2005). TEs need to strictly recognize the border between themselves and their host DNA and perform events that result in their own spread within the host genome (reviewed by Plasterk, 1995). For selfish propagation TEs encode enzymes specially dedicated to the task. Furthermore, TEs have to be able to replicate faster than the host cell but regulate their transposition to avoid self-destruction through host cell death (reviewed by Plasterk, 1995).

1.1.1 Impact on genomes

Although TEs in genomes are mostly ancient and inactivated by truncations or rearrangements, some active elements cause new insertions that are most likely benign, very occasionally harmful, and with extreme rarity beneficial (reviewed by Goodier and Kazazian, 2008; Gray, 2000). TEs have not only influenced strongly on the evolution of their host genome but also co-evolved with their hosts (Kidwell and Lisch, 1997; reviewed by Feschotte, 2008; Jurka et al., 2007). When TE insertion confers a selective benefit to the host, it is said to be domesticated. The co-option of TE sequences, their enzymatic machinery and activities by host organisms provides an additional level of variation for

organisms (reviewed by Brookfield, 2005; Feschotte and Pritham, 2007; Muotri et al., 2007; Oliver and Greene, 2009; Volff, 2006). One of the most studied example of domesticated TE enzyme is RAG1 endonuclease, which performs V(D)J recombination, a process that generates adaptive vertebrate immune system by assembling immunoglobulin and T-cell receptor genes (reviewed by Gellert, 2002). In prokaryotes, TEs may carry and spread a variety of genes involved in accessory cell functions, such as resistance to antimicrobial agents (reviewed by Bushman, 2002).

The dynamic co-evolutionary interaction between TEs and the genomes in which they reside have had a key role in the evolution of eukaryotic gene regulation (reviewed by Feschotte, 2008; Oliver and Greene, 2009; Rebollo et al., 2010; Zamudio and Bourc'his, 2010). TEs encode mechanisms for controlling their own DNA transfer and also organisms have evolved complex mechanisms to control activity of TEs (reviewed by Brookfield, 2005; Bushman, 2002; Fedoroff, 2002; Nagy and Chandler, 2004; Pritham, 2009). Therefore far beyond being junk DNA, TEs have proven to be natural genetic engineering systems and major evolutionary driving forces with important biological implications (Shapiro, 2010).

1.1.2 Classification of elements

The abundance and extreme diversity of TEs has challenged and confused their classification. However, the major distinguishing feature of TEs is whether their transposition includes an RNA intermediate (Class I) or transposition relies exclusively on DNA intermediates (Class II) (reviewed by Finnegan, 1989). These two classes of TEs can further be subdivided into separate subclasses, orders and superfamilies on the basis of the transposition mechanism, sequence similarities and structural relationships (reviewed by Wicker et al., 2007). Both classes of elements include autonomous elements that encode their own transposition machinery and non-autonomous elements that rely on autonomous elements in their transposition.

1.1.2.1 Class I elements

Class I elements use an RNA intermediate in transposition (reviewed by Boeke and Stoye, 1997; Finnegan, 1990). The genomic element is transcribed into an RNA intermediate by RNA polymerase and reverse-transcribed into DNA by an element-encoded reverse transcriptase (RT) (reviewed by Boeke and Stoye, 1997; Finnegan, 1990). Class I elements (retroelements or retrotransposons) use transposition mechanism, in which element is not excised from the donor molecule but each complete replication cycle produces complementary DNA (cDNA) copy of the element into a new target site (reviewed by Beauregard et al., 2008; Jurka et al., 2007; Pritham, 2009; Wicker et al., 2007).

Classically, retrotransposons are divided by the presence or absence of long-terminal repeats (LTRs) to LTR retrotransposons and non-LTR retrotransposons. LTR retrotransposons reverse transcribe retroviral RNA into cDNA copy within the viral particle and insert cDNA into the host chromosome involving an association of an integrase enzyme. Endogenous retroviruses residing in genomes belong to the group of LTR retrotransposons. The abundance of LTR retrotransposons is usually low in fungi, highly variable in animals, and high in plants (reviewed by Beauregard et al., 2008; Jurka et al., 2007; Pritham, 2009; Wicker et al., 2007).

Non-LTR retrotransposons often encode endonuclease, which produces a nick in the target DNA that serves as a primer for reverse transcription of their genome. Reverse transcription and integration occur through a coupled process termed target-primed reverse transcription (reviewed by (Beauregard et al., 2008). The most prevalent examples of non-LTR retrotransposons are LINEs and SINEs (long and short interspersed elements). LINEs encode proteins for their mobilization, but SINEs are nonautonomous elements and require LINEs for their propagation. Non-LTR retrotransposons are highly abundant in eukaryotes; especially in mammals (reviewed by Goodier and Kazazian, 2008).

Novel retrotransposons are identified that differ from both LTR and non-LTR retrotransposons (reviewed by Beauregard et al., 2008). One is *DIRS*-like elements (*Dictyostelium* intermediate repeat sequence) which are flanked by unusual termini and they encode a tyrosine recombinase instead of integrase enzyme (reviewed by Goodwin and Poulter, 2001; Goodwin and Poulter, 2004). The second is *Penelope*-like elements, which have LTR-like termini or are 5'-truncated (Evgen'ev et al., 1997; reviewed by Evgen'ev and Arkhipova, 2005). These elements appear to transpose via target primed reverse transcription similar to non-LTR retrotransposons (reviewed by Beauregard et al., 2008).

1.1.2.2 Class II elements

Class II elements (DNA transposons) transpose directly as DNA, utilizing either double- or single-stranded DNA intermediate (reviewed by Feschotte and Pritham, 2007; Mizuuchi, 1992b). DNA transposons are found in all kingdoms of life but are particularly prevalent in bacteria. DNA transposons are composed of specific sequences that define the ends of the element and flank an open reading frame (ORF) encoding a catalytic enzyme, the transposase (Figure 1.). Typically transposon end sequences are in inverse orientation (terminal inverted repeats, TIRs) (reviewed by Mahillon and Chandler, 1998; Mizuuchi, 1992b). For their movement, most DNA transposons use a strand exchange mechanism involving excision and reinsertion of an element, but some DNA elements use replicative transposition (chapter 1.2.2).

1.1.2.2.1 Prokaryotic elements

DNA transposons range widely in size and complexity in bacteria (Figure 1.). The simplest prokaryotic DNA elements are insertion sequences (ISs) consisting of two short TIRs flanking a single, or sometimes two, short open reading frames encoding a transposase. More genetically complicated elements are composite transposons, which are composed of a pair of IS elements that bracket additional genetic information for antibiotic resistance or other properties (for example Tn*5* and Tn*10*). Complex transposons do not have embedded ISs but may encode for several proteins (for example Tn*7*). Most intricate elements are transposing bacteriophages encoding an extensive set of proteins necessary for their viral life-style (for example phage Mu). Conjugative transposons exhibit hybrid properties of transposons, plasmids, and bacteriophages (for example Tn*916*) (reviewed by Bushman, 2002; Craig et al., 2002). The simple ISs are small; generally shorter than 2.5 kb, but complex bacteriophage Mu is even 37 kb in length.

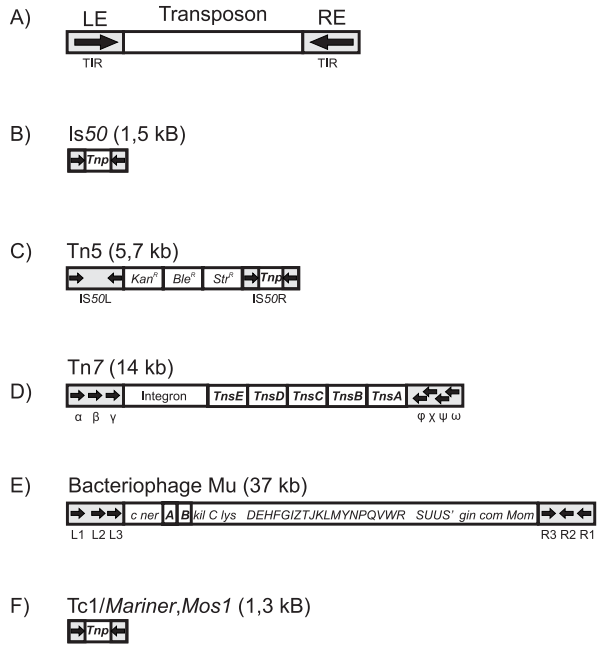

Figure 1. Examples of different DNA transposons and their end organization (not to scale). A) DNA transposons are composed of specific sequences that define the ends of the element designated to the Left End (LE) and the Right End (RE). At each transposon end (gray rectangles), there are either single inverted repeat or a mix of terminal inverted repeats (TIRs) (black arrows). Different types of prokaryotic transposons are: B) insertion sequences (ISs), C) composite transposons, D) complex transposons (in Tn*7* antibiotic resistance cassettes are part of an integron element), and E) bacteriophage using transposition (Bacteriophage Mu has two genes involved in transposition, A and B (bold)). Other genes are related to non-transposition functions (Morgan et al., 2002). F) Eukaryotic DNA transposons are similar to prokaryotic ISs. Drawn according to Craig et al., 2002.

1.1.2.2.2 Eukaryotic elements

Most DNA transposons identified in eukaryotes are quite similar to classical prokaryotic ISs (Figure 1.). DNA transposons in eukaryotes can be divided in three subclasses depending on their transposition mechanism: classic "cut-and-paste" transposons, rolling-circle mechanism utilizing *Helitrons*, and self-replicating *Mavericks* (also known as *Polintons*) (reviewed by Feschotte and Pritham, 2007; Wicker et al., 2007; Yuan and Wessler, 2011). "Cut and paste" transposons can be further subdivided, primarily on the basis of sequence similarity of the transposase, into 19 currently recognized superfamilies (see Repbase (Jurka et al., 2005); http://www.girinst.org/repbase/index.html) of which the most studied are *hAT*, *Tc1/mariner, P, PiggyBac,* and *PIF/Harbinger* elements. Class II elements include also non-autonomous derivatives called MITEs (miniature inverted-repeat transposable elements) that are composed simply of terminal IRs and a small non-coding DNA segment (reviewed by Feschotte et al., 2002).

1.1.3 Transposon tools

In addition to their basic biological interest, transposons have been used as diverse molecular tools in prokaryotes and eukaryotes during the last three decades. Transposon applications are based on harnessing the inherent features of TEs in catalyzing their own insertion into variable target sites and performing breaking and rejoining of DNA. The ability of transposons to produce random insertion mutations efficiently into various genomic sites

and potentially alter gene expression is widely exploited both at a genomic level and in the analysis of individual genes and proteins. Traditional transposon applications depended on transposition reaction *in vivo*, but the increasing understanding on transposition mechanisms promoted their *in vitro* use (reviewed by Berg and Berg, 1995; Hamer et al., 2001; Hayes, 2003; Kaiser et al., 1995; Mizuuchi and Baker, 2002). *In vivo* transposition relies either on an endogenous transposon residing in the host chromosome or introduction of plasmid containing transposon (study II). For *in vitro* applications, transposition reaction is performed with purified components in the test tube, after which the mutated DNA can be transferred into the host cells for further analysis. In their most uncomplicated format, *in vitro* transposition reaction can be reproduced in a simple reaction buffer using only transposon terminal inverted repeats, purified transposase, and the DNA target substrate (reviewed by Hayes, 2003). Combination of *in vitro* and *in vivo* approaches involves initial assembly of transposition complexes *in vitro* and delivery into cells, where transposon DNA is integrated into the genome (reviewed by Hayes, 2003). The *in vitro* conditions overcome several limitations present in traditional *in vivo* systems, e.g. host-range limitations are avoided (reviewed by Hamer et al., 2001; Mizuuchi and Baker, 2002; Reznikoff, 2008).

Transposon-based approaches are simple and highly efficient tools for genomics and proteomics studies, and are routinely used in prokaryotes, plants, and invertebrates but recently also in vertebrates including human cell lines (reviewed by Berg and Berg, 1995; Hayes, 2003; Ivics et al., 2009; Ivics and Izsvák, 2010; Kahlig et al., 2010; Kaiser et al., 1995; Ni et al., 2008). Some of the most feasible for application purposes are the bacterial systems Mu, Tn*5*, Tn*7*, Tn*10* and Tn552 as well as the eukaryotic *piggyBac* and molecularly reconstructed forms from ancient inactivated *Tc1/mariner* family elements *Sleeping Beauty* and *hAT* family element *Tol2* (reviewed by Grabundzija et al., 2010; Hayes, 2003; Ivics and Izsvák, 2010; Mizuuchi and Baker, 2002). Minimal requirements for transposon sequence are transposon ends, which contain transposase binding sites essential for mobilization of DNA between them. Therefore, for different purposes transposon can easily be engineered to carry any desired genes between transposon end sequences, e.g. different selectable marker genes, reporter genes, sequencing primer binding sites, sequences for site-specific recombination systems, controlling elements such as promoters, transcription termination signals, polyadenylation sequences, splice acceptors, or replication origins (reviewed by Berg et al., 1989; Berg and Berg, 1995).

In addition to their use as tools for analyzing functions of individual genes and their products, transposons have been used in genome-wide insertional mutagenesis projects (reviewed by Berg and Berg, 1995; Hamer et al., 2001; Hayes and Hallet, 2000; Hayes, 2003; Kaiser et al., 1995; Mizuuchi and Baker, 2002; Singh et al., 1997). Insertional mutagenesis using transposon disrupts and manipulates genes on a genome-wide scale enabling high-throughput functional studies of genes associated with various biological pathways. These types of studies have become especially valuable after completion of extensive whole-genome sequencing projects (reviewed by Judson and Mekalanos, 2000). Transposon insertions are widely applied also in other types of molecular tools, such as for sequencing strategies (Adey et al., 2010; Gertz et al., 2012; reviewed by Hayes, 2003; Mizuuchi and Baker, 2002) and various protein engineering applications (reviewed by

Hayes and Hallet, 2000; Hayes, 2003; Mizuuchi and Baker, 2002), and also to identify cancer causing genes and tumor suppressors (Copeland and Jenkins, 2010). The current developments portray transposon tools productive for gene delivery for different organisms and modification also vertebrate genomes (reviewed by Hamer et al., 2001; Ivics et al., 2009; Ivics and Izsvák, 2010; Miskey et al., 2005). Because transposon-based vectors can be used for stable genomic integration in regulated and highly efficient manner, transposons are indispensable tools for generation of transgenic cells in tissue culture and in production of germline transgenic animals for basic and applied research (reviewed by Ivics et al., 2009; Ivics and Izsvák, 2010). Transposon systems have been used for modifications of human embryonic stem cells (hESCs) and reversible production of induced pluripotent stem cells (iPSCs) (reviewed by Nieminen et al., 2010; O'Malley et al., 2009) making transposons attractive vehicles in therapeutic somatic gene transfer in humans (reviewed by Hackett et al., 2010). For these purposes, the main limitation has been the low transposition efficiency, which has been tried to overcome by hyperactive transposase constructs (Doherty et al., 2012; reviewed by Hackett et al., 2010). Furthermore, the ongoing investigations on TEs are expected to lead to the development of novel transposon technologies for genetic and cell engineering.

1.2 DDE transposons

DDE transposases are ubiquitous and represent the majority of characterised transposases, whose overall catalytic mechanism has been characterised in detail (Mizuuchi, 1992a; reviewed by Mizuuchi, 1992b). Members of the DDE transposase family carry a conserved triad of acidic residues –a DDE motif. The three acidic residues are crucial in the coordination of divalent metal ions required for catalysis (Kulkosky et al., 1992). The abundant DDE transposase family includes prokaryotic insertion sequences (ISs), members of the Tn*3* family of transposons, the Tn*7*, Tn*5* and Tn*10* families and transposable bacteriophages such as phage Mu (Nagy and Chandler, 2004, reviwed by Craig et al., 2002) and eukaryotic "cut and paste" transposons (Jurka et al., 2005; Yuan and Wessler, 2011). The family can also be extended to include retroviruses such as HIV, which encodes a catalytic integrase protein similar to the DDE transposases (Dyda et al., 1994; Haren et al., 1999; Rice et al., 1996; Rice and Baker, 2001).

Despite the vast variety of transposons, certain basic features of transposition are shared among elements, although the detailed translocation mechanisms are element-specific (reviewed by Curcio and Derbyshire, 2003; Dyda and Hickman, 2008; Hickman et al., 2010; Mizuuchi, 1992b; Turlan and Chandler, 2000). The process of transposition can be divided into a series of conceptually simple steps (reviewed by Craig, 1995; Craig et al., 2002; Grindley et al., 2006; Plasterk, 1995). In all transposons, the element-specific transposase protein recognizes and binds to the recombination sequences most commonly at the transposon termini. Transposase synapses the two element ends as a multimer and assembles into a high-order protein-DNA complex. Within the complex transposase catalyzes a set of chemically similar reactions at the ends of a transposon, the breakage and formation of phosphodiester bonds are called a donor DNA cleavage and DNA strand transfer. Subsequent reactions ultimately attach the element DNA to the target DNA. Finally, the synaptic complex is resolved and the transposition products are released.

1.2.1 Chemistry of DDE transposition

The chemistry of transposition reaction is identical for all DDE transposons. Transposition reaction proceeds through a series of single-step, in-line nucleophilic attacks resulting in two or more sequential phosphoryl transfer reactions (reviewed by Mizuuchi, 1992a; Mizuuchi and Baker, 2002; Montaňo and Rice, 2011). The nucleophilic groups used for cleavage and strand transfer are activated for catalysis by a two divalent metal ion mechanism (Nowotny et al., 2005; Nowotny and Yang, 2006; reviewed by Nowotny, 2009; Yang et al., 2006). Two divalent metal ions, which are usually $Mg2^{+}$, assist the nucleophilic attack by orienting the reacting groups, lowering the pKas of the nucleophile and leaving group, and stabilizing the geometry and charge of the transition state (Nowotny and Yang, 2006; reviewed by Mizuuchi and Baker, 2002; Montaňo and Rice, 2011). The DDE motif residues coordinate by chelation of the two assisting divalent metal ions (Kulkosky et al., 1992). The mechanism of DDE transposition reaction progresses without covalent enzyme-substrate intermediates, and it does not require external energy sources, because exchanging one high-energy phosphodiester bond for another conserves the energy.

The two chemical reactions of transposition are mechanistically similar and logically catalyzed by the same transposase active site (Kennedy et al., 2000, reviewed by Montaňo and Rice, 2011). Firstly, the transposon DNA is cleaved by hydrolysis of the 3'-terminal phosphodiester bond at each transposon end, in a nucleophilic substitution attack of OH from water on a phosphate. This donor cleavage reaction introduces a single-strand nick and exposes a free 3'-OH group at the ends of the element. Secondly, these two 3'-OH groups act as nucleophiles directly attacking the target DNA on new phosphodiester bonds at the insertion site. Concerted transposon integration involves a pair of transesterification reactions at each 3'-end of transposon DNA to one 5'-end of the target DNA. Because the two strands of the target DNA are attacked normally of few base pairs apart, the inserted transposon is flanked by short single strands of the host. The DNA intermediate generated by transposition reaction cleavage and strand transfer is converted to final recombination product by cellular enzymes. After repair replication by the host enzymes, transposon is enclosed by short duplication of the target site. The length of the target site duplication depends on the staggered cut made during strand transfer (usually 2 to 9 bp) (reviewed by Craig, 1995).

1.2.2 Transposition mechanisms of DDE transposons

In spite of their shared transposition chemistry, DDE transposons use different mechanisms for their movement to new sites. Generally, transposition strategies can be divided in two types depending on whether the donor cleavage involves single strand or double strand DNA cuts at both ends of the element (see reviewed by Hickman et al., 2010; Turlan and Chandler, 2000). Thus, the transposon either remains bound to the flanking host DNA or is completely released from the flanking DNA. The outcome results in replicative or "cut and paste" transposition (Figure 2).

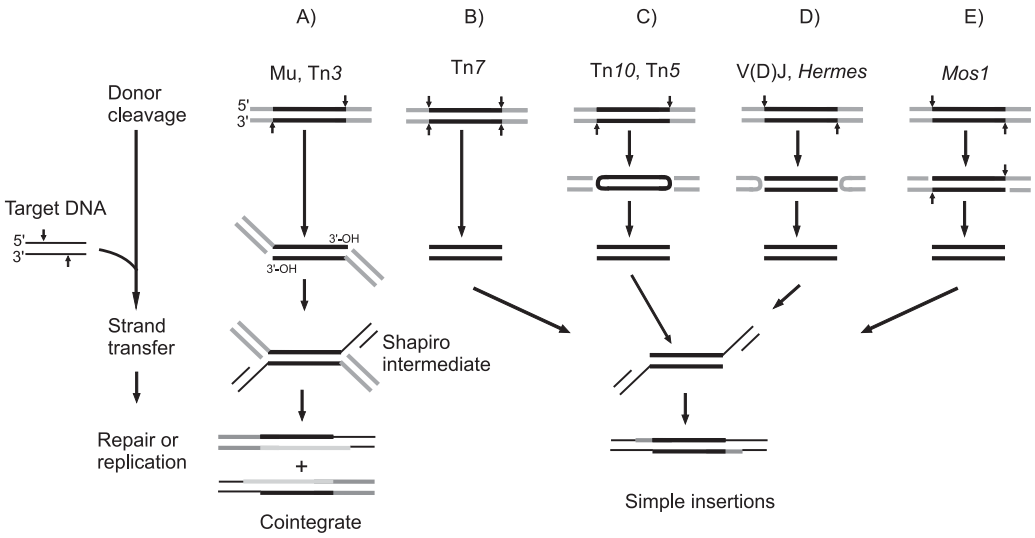

Figure 2. DDE transposition mechanisms. All transposable elements (black lines) are mobilized by two common steps, a donor cleavage (indicated by small vertical arrows) and a strand transfer to the target (thin black lines). DDE-transposons either copy themselves using replicative "copy-and-paste" mechanism (A) or use nonreplicative "cut-and-paste" mechanism for excision of a double-stranded copy of the element (B-E). In replicative transposition, only single strand cleavages are introduced to expose the reactive 3'-OH ends, which are subsequently joined to the exposed 5'-ends of the cleaved target DNA (thin black line). Replicative transposition produces a branched structure, which is replicated to yield cointegrate. In non-replicative transposition, different strategies are used for the second strand cleavage, either without hairpins (B, E) or via hairpin formation (C, D). Strand transfer is identical to all of the elements. Target cleaved in a staggered manner generates single strand gaps, which are repaired by host functions. Drawn according to Curcio and Derbyshire, 2003.

1.2.2.1 Replicative transposition

Replicative transposition is used when the donor cleavage involves cuts only at the 3'-ends of the transposon and 5'-ends of the element remain linked to the flanking host DNA (Figure 2A). Therefore, the strand transfer results in a fusion between the donor and target DNA molecules to form a branched structure, so-called Shapiro intermediate (Shapiro, 1979), in which each transposon end is joined to the donor by one strand and the target by the other. Replication of the transposon from forks created at both ends of the transposon results in formation of a composite structure called cointegrate containing both donor and target DNA joined by direct repeats of the element. Finally, cointegrate is resolved by recombination (site-specific or generalized) into the donor and recipient replicon each carrying a copy of the element. Replicative transposition can also be intramolecular that yields inversions and adjacent deletions (reviewed by Ahmed, 2009). Elements that transpose by replicative mechanisms include bacteriophage Mu (reviewed by Chaconas and Harshey, 2002) and members of the Tn*3* (reviewed by Grindley, 2002) and IS*6* families (reviewed by Chandler and Mahillon, 2002).

1.2.2.2 Cut-and-paste transposition

Most DDE transposons employ a "cut-and-paste" mechanism (or conservative transposition, non-replicative transposition) for their movement, which involves double-stranded excision of an element from the donor molecule and reinsertion into a new target site (Figure 2B-E). Although the transposition reaction itself does not involve a replication step, transposon multiplies when element is moved from replicated to non-replicated DNA during chromosome replication, or when gap repair involves homologous gene conversion from sister chromatid (reviewed by Brookfield, 1995).

"Cut-and-paste" transposons have adopted a variety of strategies for cleavage of the second strand to liberate themselves prior strand-transfer (reviewed by Hickman et al., 2010; Turlan and Chandler, 2000). Transposon Tn*7* requires two different strand-specific exonucleases for excision: TnsB catalyzes cleavage of the 3'-end and TnsA cleavage of the 5'-end (Bolland and Kleckner, 1996; Figure 2B). Members of the eukaryotic Tc1/*mariner* family transposons and related IS*630* elements cleave the non-transferred strand first at several bases within the transposon and subsequently the second strand supposedly by direct hydrolysis of the transposase itself (Feng and Colloms, 2007; reviewed by Plasterk et al., 1999; Figure 2E). Another way for second strand cleavage has been evolved for the members of IS*4* family, Tn*5* (IS*50*), Tn*10* (IS*10*), and eukaryotic *PiggyBac* superfamily (reviewed by Haniford, 2006; Reznikoff, 2008). They are excised in a three-step reaction involving a hairpin intermediate (Figure 2C). Hairpin is formed when a free 3'-OH group at cleaved transposons ends attacks the complementary strand instead of the target DNA. Hairpin formations result in the release of transposon from flanking DNA. In the third step, hairpin is hydrolyzed to regenerate the 3'-OH and 5'-phosphate ends, followed by the strand transfer to a target DNA molecule. Eukaryotic *Hermes* transposon (member of *hAT* family) has been observed to form the hairpins on the donor backbone ends instead of the excised transposon (Zhou et al., 2004; Figure 2D). Similar three-step mechanism is used to generate immunoglobin diversity in V(D)J recombination (reviewed by Gellert, 2002).

However, conservative transposition and replicative transposition are not totally two independent mechanisms but rather alternative outcomes of the transposition reaction (reviewed by Ahmed, 2009). The switch from conservative transposition to replicative transposition has been reported for Tn*5* (Ahmed, 1991; reviewed by Ahmed, 2009), Tn*7* (May and Craig, 1996), and IS*903* (Tavakoli and Derbyshire, 2001). The choice between the two alternatives is dependent whether the free 3'-OH groups engage in strand-transfer before, or after, the formation and resolution of hairpins (reviewed by Ahmed, 2009). This in turn reflects the relative efficiency of transposase to perform these two different processes at the 3'-OH ends. Thus, transposons using cut-and-paste transposition are efficient in hairpin formation. Hairpin formation also reflects the necessity of protecting free transposon ends from degradation, which is avoided by the mechanism itself in replicative transposition (reviewed by Snyder and Champness, 2007).

1.2.3 Structure of DDE transposases

Several structural and functional studies have defined topological independent functional domains for DDE transposases (reviewed by Haren et al., 1999; Hickman et al., 2010;

Polard and Chandler, 1995; Rice and Baker, 2001). Although, transposases have significant variation in size and number of the domains, their functional organization has a general pattern: sequence-specific DNA binding located at N-terminal region, the catalytic domain often localized toward the C-terminal end, and C-terminus may contain additional domains involved in a number of functions assisting transposition (reviewed by Haren et al., 1999; Hickman et al., 2010; Nesmelova and Hackett, 2010; Nowotny, 2009; Rice and Baker, 2001). The increasing number of crystal structures of catalytic cores of several transposases have been solved for both the prokaryotic (MuA; Rice and Mizuuchi, 1995 and Tn5: Davies et al., 2000) and eukaryotic Hermes (member of the eukaryotic *hAT* superfamily*:* Hickman et al., 2005) and *Mos1* (member of the eukaryotic Tc1/*mariner* family: Richardson et al., 2009) elements, which have defined the architecture of these enzymes and revealed a remarkable structural similarity of their catalytic core (Figure 3.).

The DDE residues fold into a close proximity for catalysis in a structure called an "RNase H-like fold", which was first identified in *Escherichia coli* RNase H1 (Katayanagi et al., 1990; Yang et al., 1990). Same fold is shared by all of the members in the large family of polynucleotidyl transferases (reviewed by Hickman et al., 2010; Nowotny, 2009; Rice and Baker, 2001). The conserved RNase H-like fold is a three-layered α/β/α domain (β1- β2-β3-α1-β4-α2/3-β5-α4-α5) with a central, mixed five-stranded β-sheet (reviewed by Hickman et al., 2010; Nesmelova and Hackett, 2010; Nowotny, 2009; Rice and Baker, 2001; Figure 4.). The positions of the catalytic DDE/D residues are always on the same topological elements of the fold, with the first D in the middle of the first β-strand, the second D is at the end of the fourth β-strand or just after it, and the last E/D on the fourth α-helix or just before it (reviewed by Hickman et al., 2010; Nowotny, 2009). The first three strands of the β-sheet are anti-parallel and the shorter fourth and fifth strands run parallel to the first strand. The most conserved α-helix of the fold is located after strand three and runs across one face of the β-sheet, where it stabilizes and reinforces the central β-sheet (reviewed by Hickman et al., 2010; Nowotny, 2009).

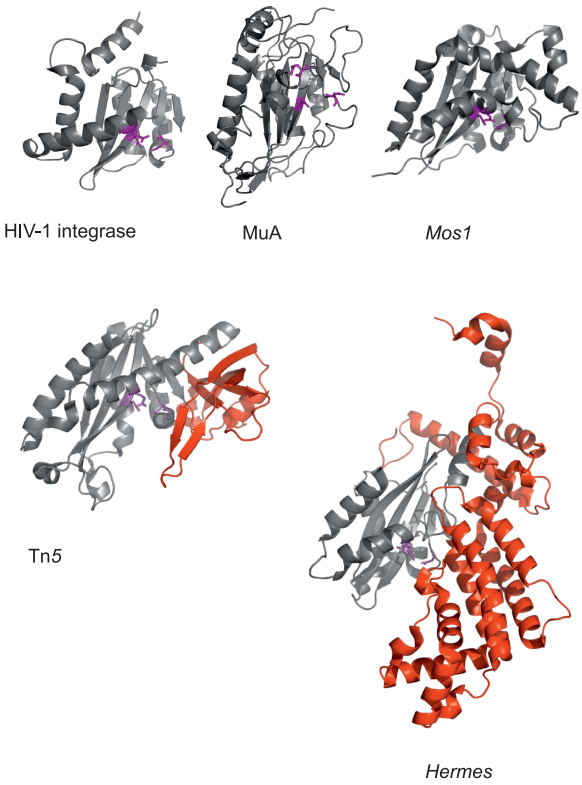

Figure 3. Ribbon diagrams of the aligned catalytic core structures of four DNA transposases and of HIV-1 integrase (PDB: 1biu, 1bco, 2fzt, 1mus, and 2bw3). The active-site residues are coloured magenta. The insertion domains of Tn5 transposase and Hermes are shown in red. The figures are rendered with PyMol.

Figure 4. Ribbon diagram of the RNase H-like fold of MuA (PDB 1bco). The central β-sheet (with numbered strands) and the conserved α-helix are shown in orange and green, respectively. More divergent parts are shown in gray. The active-site residues are coloured magenta**.** The figure is rendered in PyMOL.

The RNase H-like fold can be disrupted by insertion of sequences of various length and structure, most often to occur between β5 and α4 (reviewed by Hickman et al., 2010; Nowotny, 2009; Figure 3). This region may have additional significance in the biochemistry of transposition (reviewed by Hickman et al., 2010). In particular, large insertion domains to this region have contributed to hairpin formation and oligomerization (Davies et al., 2000; Hickman et al., 2005). Some transposases have also another sequence motif, a so called "YREK motif". It is located on α4, where the E of the YREK motif is the same as that of the catalytic DDE motif (Rezsohözy et al., 1993). It is likely that Y, R, and K form contacts important to hairpin formation on the transposon ends (Naumann and Reznikoff, 2002). However, some transposons contain only the K (or R) of the YREK motif that seems also to be catalytically important (reviewed by Hickman et al., 2010; Reznikoff, 2008).

The catalytic core of DDE transposases is not conserved beyond an RNase H-like fold, and the domains outside the catalytic domain vary considerably (reviewed by Hickman et al., 2010). However, site-specific DNA binding is typically on separate domain or domains upstream of the catalytic domain. Sequence-specific DNA binding to transposon ends is carried out by various kinds of DNA-binding domains such as one (IS*911*: Loot et al., 2002) or two helix-turn-helix (HTH) domains (Tc1/*Mariner* family: Richardson et al., 2009; van Pouderoyen et al., 1997; Watkins et al., 2004), winged-helix domains followed by an HTH domain (bacteriophage Mu: Clubb et al., 1994; Clubb et al., 1997; Schumacher et al.,1997), or a non-standard helical domains (Tn*5*: Davies et al., 2000 and *Hermes*: Hickman et al., 2005). In addition, most DDE transposases contain domains downstream of the RNase H-like catalytic domain involved in various functions such as non-specific DNA contacts, multimerization and auxiliary protein binding.

1.2.4 Transposition happens in context

1.2.4.1 Organization of transpososome

The multiple steps of transposition take place in the context of a large protein-DNA complex known as a synaptic complex or a transpososome, which contains transposon DNA ends synapsed by a multimer of transposase, and in some cases, accessory protein and DNA factors as well (Lavoie et al., 1991; Surette et al., 1987). Transpososomes are the elaborate molecular machineries of transposition, which provide a precise architecture within which the chemical reactions occur (reviewed by Gueguen et al., 2005; Hickman et al., 2010; Montaño and Rice, 2011). These higher-order complexes, which have multiple proteins bound to multiple sites, have also a high level of cooperativity in their assembly. The assembly is usually facilitated by DNA supercoiling and accessory proteins (reviewed by Chaconas and Harshey, 2002; Craigie, 1996b; Echols, 1986). The complex circuit of interactions leading to transpososome assembly varies remarkably among elements and ensures high level of specificity (reviewed by Hickman et al., 2010; Nesmelova and Hackett, 2010).

The structural and functional core of transpososome is a multimer of transposase with monomeric units symmetrically positioned at each end of the element. Dimeric assembly has been observed for e.g. Tn*5* (Davies et al., 2000), Tn10, Mos1 (Richardson et al., 2009) and the human Hsmar1 (Claeys Bouuaert et al., 2011) transposases, but MuA (Baker and Mizuuchi, 1992; Mizuuchi et al., 1992) and Mariner transposases Tc3 (Watkins et al., 2004), Sleeping Beauty (Izsvák et al., 2002), and human Himar1 (Lipkow et al., 2004) assemble into a tetramer. In addition, even hexameric assembly has been resolved for Hermes (Hickman et al., 2005). Transposition machinery of Tn*7* is exceptional as it assembles into a heteromeric complex TnsABCDE containing five separate proteins with specific functions (Waddell and Craig, 1988).

Transposases have been shown to form multimers through special multimerization domain (IS*3* family: Loot et al., 2002 and Hermes: Hickman et al., 2005), or specific sequence regions (IS911: Haren et al., 1998) or protein-protein interactions between several DNA-binding domains present in a dipartite arrangement in which two small N-terminal domains are connected by a flexible linker (Mos1: Richardson et al., 2009 and Tc3 transposase: Watkins et al., 2004). However, while some transposases form multimers by themselves, other multimerize only upon binding to transposon DNA (MuA: Baker and Mizuuchi, 1992 and Tn5: Davies et al., 2000). In addition, many transposons (e.g. Mu, Tn*10*) involve assisting host factors essential for transpososome assembly, but some transposons require also additional DNA factors (e.g. Mu) or presence of target DNA (e.g. Tn*7*) (reviewed by Chaconas and Harshey, 2002; Craig, 2002; Reznikoff, 2002). Therefore, formation and stabilization of synaptic complexes is due to contribution of intertwined network of protein-DNA and protein-protein interactions of various protein and DNA components in the transpososome organization (Davies et al., 2000; Hare et al., 2010a; Montaño et al., 2012; Richardson et al., 2009, reviewed by Hickman et al., 2010; Montaño and Rice, 2011).

1.2.4.2 Function of transposition machineries

Transpososomes are dynamic entities, which undergo staged conformational changes to accommodate the different steps along the transposition pathway (reviewed by Gueguen et al., 2005). The increasing stability of the transpososome as it evolves to a productive complex pulls transposition forward together in addition to product binding energy (Abdelhakim et al., 2010; Montaňo et al., 2012; Yanagihara and Mizuuchi, 2003; reviewed by Chaconas and Harshey, 2002). The specialized nucleoprotein structures regulate biochemical reactions with high precision (reviewed by Craigie, 1996b; Echols, 1986). The particular arrangement of DNA and protein components in the transpososome is prerequisite for assembly and catalysis but transposition machineries also regulate frequency, precision, directionality, and mechanism of transposition (reviewed by Harshey and Jayaram, 2006; Yin et al., 2007). Control of activity is important to avoid undesirable reactions and assure coordination of the multiple distinct transposition reaction steps catalyzed by the same active site, which occupies sequentially the different phosphate groups (Kennedy et al., 2000; reviewed by Gueguen et al., 2005; Mizuuchi and Baker, 2002; Montaňo and Rice, 2011). Recurring theme in the function of transposition machinery is that catalysis of transposition occurs in *trans*: the transposase subunit bound to one transposon end in the transpososome catalyzes the chemical steps on the other end of the transposon (Aldaz et al., 1996; Davies et al., 2000; Hare et al., 2010a; Montaňo et al., 2012; Namgoong and Harshey, 1998; Richardson et al., 2009; Savilahti and Mizuuchi, 1996). Therefore, the transpososome assembly is a crucial regulatory checkpoint of transposition reaction that retains transposase inactive until it is incorporated into a productive synaptic complex (reviewed by Craigie, 1996a; Mizuuchi et al., 1992).

The mechanistic differences among diverse transposons are mediated by or through the transposition machineries (reviewed by Gueguen et al., 2005; Mizuuchi, 1992b). Structural features and accessory proteins of the transpososome sense and exert control in target DNA capture, target immunity, and resolution of transposition products (reviewed by Gueguen et al., 2005; Parks et al., 2009). In target engagement, transpososome assess new DNA molecules for potential insertion (reviewed by Craig, 1997). Most transpososomes choose their integration sites at random, but show some sequence level preference (as for Tn*10*/IS*10*: Bender and Kleckner, 1992; Halling and Kleckner, 1982 and Mu: Haapa-Paananen et al., 2002; Haapa-Paananen et al., 2002). An exception is Tc1*/mariner* elements, which all integrate into the sequence TA (van Luenen and Plasterk, 1994). However, transposon insertions are directed into DNA sequence that can form particular conformation. For example, important in target DNA selection is target DNA bending for Tn*10* (Pribil and Haniford, 2003), a triple-helical DNA structure for Tn*7* (Kuduvalli et al., 2001; Rao et al., 2000) and DNA mismatches for Mu (Yanagihara and Mizuuchi, 2002). Additionally, replication fork has been shown as a preferred target for insertions (e.g. Tn*7*: Parks et al., 2009 and IS*903*: Hu and Derbyshire, 1998) and the nucleosome-free DNA region in eukaryotic cells (e.g. Hermes: Gangadharan et al., 2010, Mu: Liu et al., 2009; Paatero et al., 2008). Target site selection can be mediated by interactions with host factors (e.g. DNA replication processivity factor for Tn*7:* Parks et al., 2009) or different target selection proteins that require ATP to choose target DNA (e.g. TnsD and TnsE proteins for Tn7: Craig,

2002; Peters and Craig, 2001) and MuB protein for Mu (Adzuma and Mizuuchi, 1988) or using transposase itself (e.g. Tn10: Junop and Haniford, 1997; Sakai and Kleckner, 1997). In addition, some transposition systems prevent from insertion into itself or into close vicinity by the phenomenon called target immunity. In the case of Mu and Tn*7*, transposase (i.e., MuA and TnsAB, respectively) and ATP-dependent DNA binding protein involved in target capture (i.e., MuB and TnC, repectively) act together to perform target immunity (reviewed by Chaconas and Harshey, 2002; Craig, 2002).

Consequently, the transpososome architecture provides a key to understand all of the attributes of transposition, how it happens and how it is regulated (reviewed by Gueguen et al., 2005; Harshey and Jayaram, 2006; Hickman et al., 2010; Montaňo and Rice, 2011). Therefore, the three-dimensional relationships between the components of the transposition machinery and changes in their configuration reveal important aspects of the transposition reaction (reviewed by Hickman et al., 2010; Montaňo and Rice, 2011).

1.2.4.3 Transpososome structures

Despite the early success with crystallization of the catalytic cores and Tn*5* transpososome structure (Davies et al., 2000), the detailed structure of transposition machineries for other elements remained elusive for a long time (reviewed by Harshey, 2012; Hickman et al., 2010; Montaňo and Rice, 2011). The few DDE transposases, which have been structurally characterized as complexes with their DNA substrates, have provided important details of their mechanism of action. Comparison of the known three-dimensional structures of transpososomes has highlighted conservation of catalytic domains but

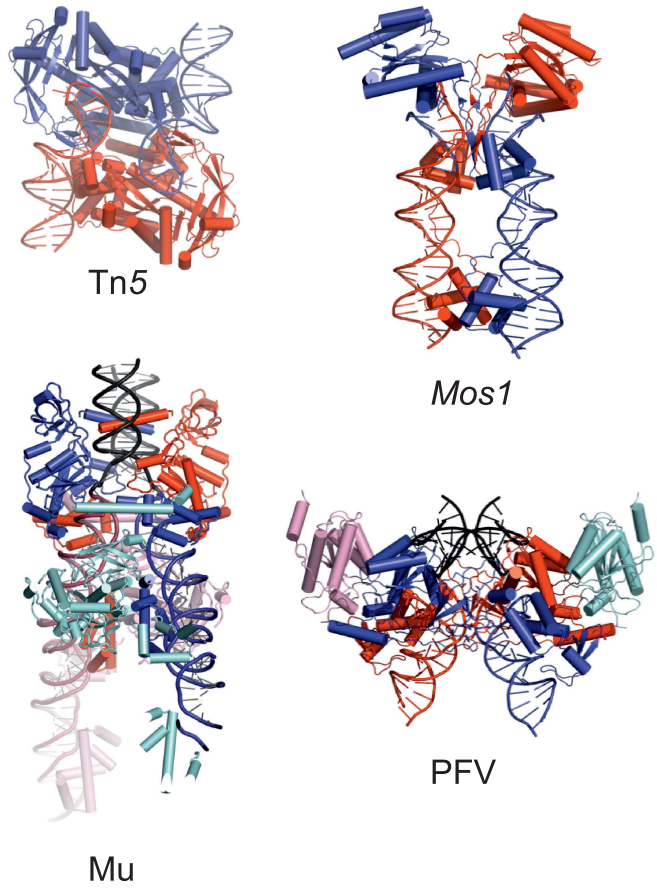

Figure 5. DDE recombinase – DNA complexes (PDB: 1mus, 2fzt, 4fcy, and 3oso). DNA ends are red and blue, and target DNA (where included) black. Catalytic subunits are red and blue. For tetrameric complexes (Mu and PFV) other subunits are shown, pink and cyan. In the PFV intasome structure, only catalytic domain of the additional subunit was visible. The figures are rendered in PyMOL.

diversity of architectures (reviewed by Dyda and Hickman, 2008; Harshey, 2012; Montaňo and Rice, 2011; Figure 5). The crystal structures of prokaryotic Tn*5* transpososome representing a synaptic complex at the stage following cleavage from donor DNA (Davies et al., 2000) and eukaryotic Mos1 transpososome poised for target capture and strand transfer (Richardson et al., 2009) were the first to open the detailed structural perspectives. The recently resolved crystal structure of Mu transposase in post-integration stage sheds more light on the DDE recombinase evolution (Montaňo et al., 2012). Furthermore, because the DDE family can be extended to include retroviruses, the crystal structures of prototype foamy virus (PVF) integrase in complex with its cognate viral DNA (Maertens et al., 2010) have gained mechanistic insights to functional complex organization in general.

The overall architectures of the Tn*5*, Mos1, and Mu transpososomes and PFV intasome are remarkably different but they share several recurring features, which may reflect convergent evolution for functional reasons (Montaňo et al., 2012; reviewed by Montaňo and Rice, 2011). The overall architecture of Tn*5* is a compact, approximately globular assembly, which contains transposase dimer bound to the two anti-parallel transposon ends (Davies et al., 2000; reviewed by Nesmelova and Hackett, 2010). The overall architecture of Mos1 looks very different with a dimer of transposase extended to transposon DNA end sequences and bound by DNA-binding domains of transposase monomers in a parallel fashion (Richardson et al., 2009). The Mu transpososome and the PFV intasome contain a tetramer of recombinase and two parallel element DNA ends (Hare et al., 2010b; Maertens et al., 2010; reviewed by Cherepanov et al., 2011). These tetramers contain two functionally distinct pairs of protomer subunits (Montaňo et al., 2012; reviewed by Cherepanov et al., 2011). The inner subunits of the tetramer are responsible for catalysis and the outer subunits seem to provide supporting functions (Hare et al., 2010b; Montaňo et al., 2012; reviewed by Cherepanov et al., 2011).

Outside of the fundamental building block, RNase H-like fold, the structures of transposases vary greatly (reviewed by Hickman et al., 2010). Accordingly, the transpososomes structures are conserved only through their catalytic cores. All transpososomes are arranged to perform catalysis *in trans*, which ensures coordinated reactions at the element ends. Also the positions of the catalytic carboxylates and metal ions and the 3'-nucleotides within the active sites are conserved despite the divergence of the architectures (reviewed by Cherepanov et al., 2011). Other similarity in the transpososome architectures is the way complexes are maintained together via intertwined network of protein-protein and protein-DNA contacts, although underlining interaction networks are mediated by different domains (Davies et al., 2000; Montaňo et al., 2012; Richardson et al., 2009; reviewed by Cherepanov et al., 2011). In transpososomes, transposon ends are bound bind by the HTH-motifs of the N-terminal binding domain (Tn5) or bipartite DNA-binding domains (Mos1 and Mu), but they form divergent contacts in their respective transpososomes (Montaňo et al., 2012; Richardson et al., 2009; reviewed by Montaňo and Rice, 2011; Steiniger-White et al., 2004). Also the synaptic complex formation is mediated by multimerization using different protein segments and a divergent network of contacts (reviewed by Cherepanov et al., 2011; Dyda and Hickman, 2008; Montaňo and Rice, 2011; Steiniger-White et al., 2004).

Another common feature to transpososomes is target bending (Montaňo et al., 2012; reviewed by Montaňo and Rice, 2011). This is revealed by complexes containing the target (Maertens et al., 2010; Montaňo et al., 2012). Furthermore, the target DNA can be modeled also onto the structures of Tn5 and Mos1 were it is presumed to adopt a bent conformation (Davies et al., 2000; Richardson et al., 2009). This feature is proposed to contribute to the irreversibility of the strand transfer reaction as the DNA is strained away from the active site after the reaction (Montaňo et al., 2012). This may be the basis for the product binding energy that pulls transposition reaction forward (Montaňo et al., 2012).

Comparison of the several synaptic complex structures has highlighted the divergent overall architectures, which reflects the structural differences of transposases and consequently diverse transposition mechanisms (replicative transposition for Mu, hairpin formation for Tn5 but not for Mos1), different spacing between attacks on the target DNA (5-bp for Mu, 9-bp for Tn5, and 2-bp for Mos1) and different regulatory details but convergent evolution for function (Montaňo et al., 2012; reviewed by Hickman et al., 2010; Montaňo and Rice, 2011; Nesmelova and Hackett, 2010).

1.3 Bacteriophage Mu

Bacteriophage Mu is both a transposable element and a temperate phage of *Escherichia coli* and other Gram-negative bacteria with the capacity to induce high rates of mutations, hence its name Mu (short for mutator) (Taylor, 1963). Mu is unique as a transposon and a phage, as while it uses two modes of transposition efficiently during the distinct stages of its life cycle: (i) to integrate its genome as a simple insert into the host DNA following infection and (ii) to replicate itself during the lytic growth of the host using replicative transposition (reviewed by Mizuuchi and Craigie, 1986). Although the initial Mu integration into the host genome results in a simple insertion (Akroyd and Symonds, 1983; Chaconas et al., 1983; Harshey, 1984; Liebart et al., 1982), it is supposedly mainly an unusual outcome of the replicative transposition mechanism due to the differences in processing the transposition intermediates (Au et al., 2006; Choi and Harshey, 2010; Jang et al., 2012).

Mu is with its 36,717-bp genome and 55 genes one of the most complex and elaborate transposition systems studied to date (Morgan et al., 2002; reviewed by Harshey and Jayaram, 2006; Figure 1E). However, the highly efficient transposition (Chaconas et al., 1981) and development of the first defined *in vitro* transposition system of Mu (Craigie and Mizuuchi, 1985; Craigie et al., 1985; Mizuuchi, 1983) has made the Mu transposition a thoroughly characterized system and a paradigm of DNA transposition in general (reviewed by Chaconas and Harshey, 2002; Harshey, 2012; Mizuuchi, 1992b) even though Mu uses replicative mode of transposition, which is more of an exception than rule among transposons. The replicative Mu DNA transposition proceeds through a series of higher-order nucleoprotein complexes termed Mu transpososome, which contain as a structural core a tetramer of Mu encoded transposase, MuA, bound to the two transposon ends (Lavoie et al., 1991; Surette et al., 1987). The assembly and function of Mu transposition machinery is a sophisticated process with an arsenal of *cis* and *trans* requirements for transposition (Baker and Mizuuchi, 1992; Craigie et al., 1985; Mizuuchi and Mizuuchi, 1993; Savilahti

et al., 1995; Yin et al., 2007). The chemical steps of transposition are catalyzed by MuA transposase, but other Mu encoded proteins, a number and intricate arrangement of phage encoded DNA sites, several host encoded proteins, and strict DNA topology requirements are involved in transposition (reviewed by Chaconas and Harshey, 2002; Harshey and Jayaram, 2006).

1.3.1 DNA requirements in Mu transposition

The DNA structure at the ends of Mu genome is asymmetric with three MuA binding sites at each end (Craigie et al., 1984; Groenen and van de Putte, 1986; Zou et al., 1991; Figure 6A). The sparsely spaced binding sites at left end, designated L1, L2, and L3 are all in the same head-to-head orientation. The ~80 bp space between L1 and L2 contains a binding site for the DNA-bending protein HU (Lavoie et al., 1996). In contrast, the binding sites at the right end are closely spaced with R3 site in the opposite orientation of R1 and R2. The binding sites share a 22-bp consensus sequence, with no obvious internal symmetry (Craigie et al., 1984; Zou et al., 1991). All six Mu end binding sites are required for optimal transposition but their relative importance for transposition varies as their MuA binding affinity (Allison and Chaconas, 1992; Craigie et al., 1984; Kuo et al., 1991; Lavoie et al., 1991; Mizuuchi et al., 1991). In addition to binding sites in transposon ends, a supplementary binding region for MuA and host cell proteins exists about 1 kb from the left end (Figure 6A).

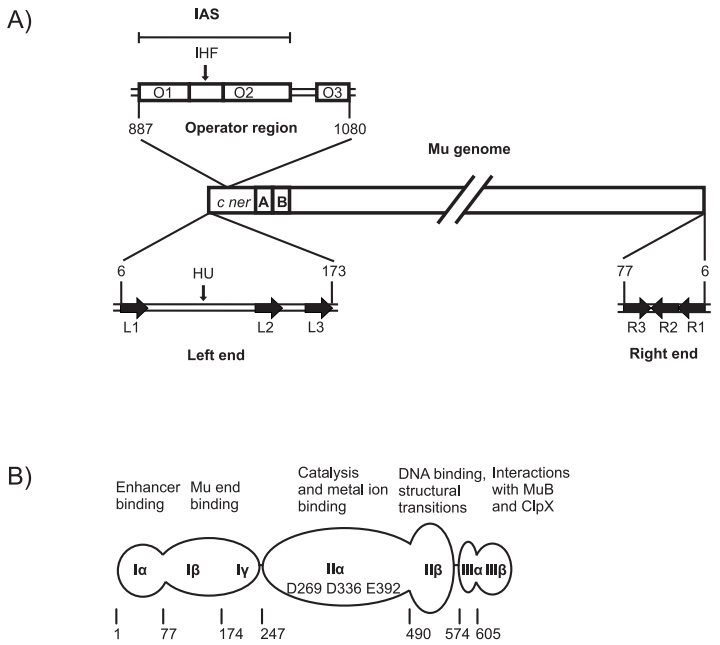

Figure 6. A) *Cis* and *trans* requirements in Mu transposition. Shown are genes *A* and *B* (bold) encoding MuA transposase and MuB protein, respectively. The substructures of the Mu left (L) end, internal activation sequence (IAS) i.e. enhancer region, and right (R) end are enlarged. Black vertical arrows denote the MuA binding sites L1 to L3 and R1 to R3, and the orientation and spacing of the consensus sequences within the sites, locations of L-end and R-end are numbered. Binding site for *E. coli* bending protein HU in L-end is shown. The upper enlargement show operator region (O1-O3), the location in L-end is numbered. It contains binding site for *E. coli* integration host factor (IHF). Drawn according to Morgan et al., 2001, Akhverdyan et al., 2011. B) Structural organization of MuA with different functions assigned to various subdomains (Roman Numerals ans Greek letters). The numbers correspond to the amino terminus of each subdomain as specified earlier *(*Schumacher et al., 1997).

Transpositional enhancer (E; O_1-O_3) or the internal activating sequence (IAS) is part of the operator region, which contains clusters of MuA-binding sequences that also bind the Mu repressor and between O_1 and O_2, a binding site for integration host factor (IHF) (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989; Surette et al., 1989). The IAS promotes authentic transpososome assembly (Mizuuchi et al., 1992; Surette and Chaconas, 1992) and increases the efficiency of transposition (Leung et al., 1989; Surette et al., 1989). Also the terminal dinucleotide 5'-TG…CA-3' of Mu is essential for stable transpososome assembly and transposition reaction (Burlingame et al., 1986; Goldhaber-Gordon et al., 2002a; Goldhaber-Gordon et al., 2002b; Lee and Harshey, 2001; Lee and Harshey, 2003a; Lee and Harshey, 2003b; Watson and Chaconas, 1996). The terminal dinucleotide is conserved among DNA transposons and assist DNA opening at the transposon termini prior transposition chemistry by allowing conformational flexibility (Lee and Harshey, 2001; Lee and Harshey, 2003a; Lee and Harshey, 2003b).

In addition to the three separate DNA sites, the Mu transpososome assembly requires negative supercoiling of Mu DNA to induce important conformational and torsional effects that are prerequisite for the specific geometry of the DNA segments within the productive synaptic complex (Mizuuchi and Craigie, 1986; Craigie and Mizuuchi, 1985; Craigie and Mizuuchi, 1986; reviewed by Chaconas and Harshey, 2002). DNA supercoiling assist specific MuA binding to end binding sites, but also binding of IHF to the IAS and HU to the spacer region between L1 and L2 (Kobryn et al., 1999; Surette et al., 1989). These accessor protein bindings induce precisely positioned DNA bends that brings distant MuA binding sites together to be linked by an oligomer of MuA (Lavoie et al., 1996). Additionally, the free energy of DNA supercoiling drives the conformational changes required for synapsis in general (reviewed by Chaconas and Harshey, 2002; Mizuuchi and Baker, 2002).

1.3.2 Organization of catalytic MuA transposase

MuA transposase is the catalytic enzyme of Mu transposon responsible for conducting all the chemical steps of transposition reaction. MuA is a 75-kDa protein (663 amino acids) and can be divided into structurally and functionally defined three major domains and several subdomains (Nakayama et al., 1987; Figure 6B). Structures of the five of its subdomains have been determined either by X-ray crystallography or by NMR (Clubb et al., 1994; Clubb et al.,997; 271 Schumacher et al., 1997, Rice et al., 1995) and recent Mu transpososome crystal structure reveals the architecture for nearly full-length MuA (Montaňo et al., 2012; chapter 1.3.5.1 and Figure 5 and 7). The N-terminal DNA binding domain I participates in transposon end and enhancer DNA binding through separate regions. The small N-terminal subdomain Iα contains a winged-HTH motif (Clubb et al., 1994) to bind the IAS and promote transpososome assembly (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989). The specific Mu end-binding is mediated through two larger subdomains Iβ and Iγ both containing HTH motif, which recognize their own subsites in the ~50-bp MuA binding site (Clubb et al., 1997; Schumacher et al., 1997). Subdomains Iβ and Iγ of the monomeric MuA bind together over the two adjacent major grooves of a MuA binding site by using their respective recognition helices of a HTH motif and interacting with the intervening minor groove on the same face on the DNA with the linker that connects Iβ and Iγ (Montaňo et al., 2012; Schumacher et al., 1997; Zou et al., 1991).

Following DNA-binding modules is the central domain II, which contains the catalytic subdomain IIα associated at its C-terminus with a subdomain IIβ (Rice and Mizuuchi, 1995). The catalytic subdomain IIα contains the critical DDE-motif of acidic residues, at positions D269, D336 and E392, which are involved in the metal ion coordination during the catalysis (Baker and Luo, 1994; Krementsova et al., 1998). The divalent metal ion used in Mu reactions is Mg^{2+}, although it can be substituted with several other cations in the distinct steps of the transposition pathway (Baker and Luo, 1994; Kim et al., 1995; Mizuuchi et al., 1992; Wang et al., 1996). Subdomain IIα forms a mixed α/β domain and contain the RNase H-like fold: a central five-stranded mixed parallel and antiparallel β sheet, with helices on either side (reviewed by Rice and Baker, 2001; Figure 4). DDE motif residues of D269 and D336 are located on adjacent strands of the β sheet, and the third, E392, is on a mobile loop that passes across the front of the sheet (reviewed by Rice and Baker, 2001). This configuration of the E392 detected in the separate domain structure (Rice et al., 1995) is unable to coordinate two metal ion cofactors (reviewed by Hickman et al., 2010; Rice and Baker, 2001). It is suggested that this disordered loop become ordered upon DNA binding and fold into a regular α-helical structure and become an upstream extension of α4 (Montaňo et al., 2012; reviewed by Hickman et al., 2010).

The subdomain IIβ forms a small, β-barrel with a positively charged surface with non-specific DNA-binding activity (Nakayama et al., 1987; Rice and Baker, 2001). The involvement of subdomain IIβ in target and Mu DNA binding was evidenced in transpososome crystal structure (Montaňo et al., 2012). Other functions have also been assigned for subdomain IIβ such as assisting in complex assembly and structural transitions (Krementsova et al., 1998; Montaňo et al., 2012; Namgoong et al., 1994).

The C-terminal domain III can be divided into two subdomains. The positively charged subdomain IIIα was structurally uncharacterized before the resolved Mu transpososome structure, which depicted a long helical structure for subdomain IIIα (Montaňo et al., 2012). Subdomain IIIα participates in target DNA capture and non-specific DNA binding and structural transitions as subdomain IIβ (Baker et al., 1993; Krementsova et al., 1998; Montaňo et al., 2012). Subdomain IIIα also displays cryptic endonuclease activity and may be associated with DNA-binding in Mu-host junction and nontransferred strand through the positively charged residues RRRQK, at positions 575-579. This region is proposed to remove the attached host DNA following the initial integration or deliver the uncleaved strand into active site for hydrolysis (Choi and Harshey, 2010; Montaňo et al., 2012; Wu and Chaconas, 1995)

The C-terminal subdomain IIIβ is responsible for interactions with other proteins that are regulators of transposition. Subdomain IIIβ interacts with the phage-encoded MuB protein, a non-specific DNA binding protein with ATPase activity, which is important in targeting transposition into distal target sites (Baker et al., 1991; Hashey and Cuneo, 1986; Leung and Harshey, 1991; Wu and Chaconas, 1992). The carboxyl terminus of subdomain IIIβ contains a binding tag for the host-encoded ClpX protein, which remodels the transpososome for disassembly before replication repair can occur (Levchenko et al., 1995).

All MuA subdomains are required for efficient phage Mu transposition *in vivo* but under some circumstances subdomains Iα and IIIβ can be excluded. These include appropriately

altered DNA substrates and/or suitably modified reaction milieu (Jiang and Harshey, 2001; Kim and Morrison, 2009; Krementsova et al., 1998; Mizuuchi and Mizuuchi, 1989; Yang et al., 1995). In the absence of enhancer, domain Iα is even known to be inhibitory (Mizuuchi and Mizuuchi, 1989; Yang et al., 1995).

1.3.3 Steps and nucleoprotein complexes in Mu transposition

Mu transposition starts with the binding of inert MuA monomers via their Iβγ domains into three sites at each end of the Mu genome, which elicits the stepwise assembly of Mu transpososome containing MuA tetramer as a structural and functional core (Craigie et al., 1984; Kuo et al., 1991; Zou et al., 1991; reviewed by Mizuuchi, 1992b). Initially, the binding of inert MuA monomers causes DNA bending at each site. Also the assisting host proteins (IHF, HU) induce precisely positioned DNA bends to promote the interactions among three DNA sites, the left (L) and right (R) ends of Mu, and enhancer element (E), present on supercoiled DNA (Kuo et al., 1991; Watson and Chaconas, 1996). This complex circuit of interactions through topological filter leads to bridging interactions with the enhancer via Iα domains, initially E-R (enhancer – right end) crossings, which progress by DNA slithering to a transient three-site-synaptic complex LER, (left end – enhancer – right end) trapping five DNA supercoils (Pathania et al., 2002; Yin et al., 2007; reviewed by Harshey and Jayaram, 2006). The DNA-protein and protein-protein interactions in LER activate tetramerization of MuA (Lavoie et al., 1991; Pathania et al., 2002; Pathania et al., 2003; Yin et al., 2005; Yin et al., 2007; reviewed by Harshey and Jayaram, 2006). Subsequently, the ordered pathway of transposition proceeds through formation of distinct tetrameric complexes with different configuration and of increasing stability (Pathania et al., 2002, reviewed by Chaconas and Harshey, 2002; Figure 6). The enhancer remains weakly associated with the MuA also in tetrameric complexes, although it is not required for chemical steps (Surette and Chaconas, 1992; Yin et al., 2005; Yin et al., 2007). A stable core of tetramer retains two L-R and one R-E DNA crossing (Yin et al., 2005; Yin et al., 2007, reviewed by Harshey, 2012).

In the first tetrameric transposition complex the transposon ends are bridged to form the stable synaptic complex (SSC or type 0 complex) with all six MuA-binding sites occupied but only three sites (R1, R2 and L1) tightly bound (Kuo et al., 1991; Lavoie et al., 1991; Mizuuchi et al., 1991). The MuA binding is extended into host DNA and DNA helix is opened at the transposon termini, which is the rate-limiting step of the cleavage reaction (Lavoie et al., 1991; Lee and Harshey, 2003a; Wang and Harshey, 1994; Wang et al., 1996). Within the tetramer, only the two MuA subunits located near the transposon termini (L1, R1) conduct the catalytic steps, whereas other two MuA subunits (bound to L2, R2) are essential for structural and functional integrity of the transpososome (Aldaz et al., 1996; Baker et al., 1993; Mariconda et al., 2000; Namgoong and Harshey, 1998; Savilahti et al., 1995; Savilahti and Mizuuchi, 1996; Williams et al., 1999; Yang et al., 1995). The terminal nucleotides are engaged into the active site (Lee and Harshey, 2003a; Savilahti et al., 1995), which is organized to perform catalysis *in trans* (L1-bound subunit promote catalysis in R1 and vice versa) (Aldaz et al., 1996; Namgoong and Harshey, 1998; Savilahti and Mizuuchi, 1996). The domain sharing between monomers in the two active sites results in the interwoven architecture of the transpososome (Montaňo et al., 2012; Yang et al., 1995; Yang et al., 1996; Yuan et al., 2005a; reviewed by Harshey and Jayaram, 2006).

In the presence of Mg^{2+}, the type 0 complex is quickly converted to type 1 complex or cleaved donor complex (CDC) as MuA subunits in L1 and R1 hydrolytically cleave the opposite Mu DNA ends at the terminal 3'-CA dinucleotides (Craigie and Mizuuchi, 1987; Mizuuchi, 1984; Surette et al., 1987). The exposed 3'-OH groups at the transposon ends are held in the active site, whose conformational changes upon target capture align the cleaved ends for strand transfer to occur. Target is delivered to the transpososome by an ATP-dependent DNA-binding protein MuB through the interaction of C-terminus of MuA with an ATP-dependent DNA-binding protein MuB, which assembles into oligomeric clusters on DNA making MuB-bound DNA as an efficient transposition target (Baker et al., 1991; Han and Mizuuchi, 2010; Surette and Chaconas, 1991). Mu maximizes transposition potential by allowing target capture at varying steps of the transposition pathway, not only at the donor cleavage step (reviewed by Chaconas and Harshey, 2002).

Short-lived type 1 complex is converted to type 2 complex or strand transfer complex (STC) when nucleophilic 3'-OH groups at transposon ends attack the phosphodiester bonds on target DNA of 5 bp apart, allowing simultaneous cleavage and joining through transesterification reaction (Mizuuchi and Adzuma, 1991). The most stable of the Mu transpososome is the type 2 complex, which contains the branched θ structure or Shapiro-type intermediate that will serve as a template for replication (Surette et al., 1987). However, this complex has to be actively disassembled before replication can occur. The complex is destabilized by unfolding of MuA subunit (either at L1 or R1) using host remodeling machine ClpX that interacts with the C-terminus of MuA (Abdelhakim et al., 2008; Abdelhakim et al., 2010; Jones et al., 1998; Kruklitis and Nakai, 1994; Kruklitis et al., 1996; Levchenko et al., 1995; Levchenko et al., 1997). The essential remodeling results in the formation of a fragile type 3 complex or STC2 complex that allows the recruitment of the host replication apparatus to resolve the branched strand transfer structure by target-primed replication, which results in formation of a cointegrate (Nakai et al., 2001). Ligation of the single-stranded gaps at the Mu-target junction by host factors completes a 5-bp duplication of the target site, considered as the hallmark of transposition (Allet, 1979; Kahmann and Kamp, 1979).

1.3.4 Control of transposition

Without control, transposition would be detrimental to the host cell, and thus for the element itself (reviewed by Plasterk, 1995). Especially, while Mu replicates itself 100-fold during the lytic phase of its life cycle, it faces also a serious threat of self-destruction by inserting into itself (Bukhari, 1975). For its own good, Mu is one of the highly regulated transposon systems studied to date (reviewed by Chaconas and Harshey, 2002; Craigie, 1996a; Harshey and Jayaram, 2006). The Mu transposition is controlled at several levels. One way of control is the ordered interaction of the enhancer (ER, LER), which regulates transposition by specifying the architecture of MuA tetramer and DNA topology required for function of the transpososome (Yin et al., 2007). The enhancer has a role as a silencer when bound by lysogenic repressor c, which is homologuous to subdomain Iα of MuA (Harshey et al., 1985). Repressor binding prevents MuA binding and thus transpososome formation (Mizuuchi and Mizuuchi, 1989). Vice versa, the involvement of enhancer as a part of the transpososome

prevents repressor binding and ensures commitment of transpososome to transposition (reviewed by Harshey and Jayaram, 2006). Repressor also negatively regulates transcription of the early gene products essential for the lytic growth, including MuA and MuB (Krause and Higgins, 1986).

Another way of control can be executed at the level of target site selection and avoiding Mu insertion into itself by a process called target immunity (or *cis*-immunity), both of which are mechanisms involving MuB, ATP and MuA (reviewed by Chaconas and Harshey, 2002). At regional level, Mu integration is targeted to DNA bound by MuB protein, which has a modest preference for A/T-rich sequence (Manna et al., 2001; Maxwell et al., 1987; Mizuuchi and Mizuuchi, 1993; Naigamwalla and Chaconas, 1997). However, at sequence level Mu integrates essentially randomly, although some target sequence preference is statistically observed (5'-C-Py-G/C-Pu-G-3') (Butterfield et al., 2002; Haapa-Paananen et al., 2002; Mizuuchi and Mizuuchi, 1993). In addition, certain DNA deformations such as single-nucleotide mismatch are identified as a hot spot for Mu integration (Yanagihara and Mizuuchi, 2002).

In target immunity, MuB-ATP is bound to DNA at high-affinity and interacts with Mu end-bound MuA (Greene and Mizuuchi, 2002a; Greene and Mizuuchi, 2002b; Maxwell et al., 1987). This interaction is achieved by the formation of DNA loops between the MuA- and the MuB-bound DNA sites (Adzuma and Mizuuchi, 1989; Greene and Mizuuchi, 2002c; Han and Mizuuchi, 2010). This interaction stimulates MuB ATPase activity resulting MuB to dissociate from the DNA (Mizuuchi and Mizuuchi, 1993). As a consequence, MuB removed from close vicinity of MuA accumulates on DNA far away from the Mu genome DNA and stimulates MuA to integrate Mu DNA to this selected distal site (10-15 kb away) (Schweidenback and Baker, 2008). However, an additional genome-immunity mechanism functional outside Mu ends has been proposed, where MuB is removed and bound strongly (Ge et al., 2010). In addition to these, MuB is also involved in transposition regulation by protecting transpososomes from a premature disassembly by the host chaperone ClpX (Levchenko et al., 1995).

1.3.5 Mu transposition *in vitro*

The establishment of Mu *in vitro* transposition system method (Craigie and Mizuuchi, 1985; Craigie et al., 1985; Mizuuchi, 1983) has been essential for the comprehensive understanding of the Mu transposition. The initial *in vitro* reaction was performed with a supercoiled donor plasmid DNA carrying transposon ends in correct orientation, replication-competent *E. coli* extracts including host factors, and extracts containing MuA and MuB proteins (Mizuuchi, 1983). Subsequently, the *in vitro* transposition reaction was refined further to contain purified protein components, ATP, and divalent cation, which allow a more detailed dissection of reaction process (Craigie and Mizuuchi, 1985; Craigie et al., 1985).

Finally, the relaxed topological requirements were permitted by addition of dimetylsulfoxide (DMSO) or glycerol either alone or in combination into the transposition reaction (Baker and Mizuuchi, 1992; Craigie and Mizuuchi, 1986; Mizuuchi and Mizuuchi, 1989). This inclusion relaxed the need for DNA supercoiling and correct relative orientation of the Mu ends, as well as the presence of enhancer (Baker and Mizuuchi, 1992; Craigie

and Mizuuchi, 1986; Mizuuchi and Mizuuchi, 1989). Under these altered conditions also HU, IHF and MuB proteins and the enhancer-binding domain Iα and the MuB-interacting domain IIIβ of MuA could be omitted (Baker and Mizuuchi, 1992; Craigie and Mizuuchi, 1986; Mizuuchi and Mizuuchi, 1989). Reaction pathway could be simplified further by using precleaved donor DNA substrates, which allowed the efficient strand transfer to a target DNA but bypassed the cleavage step (Craigie and Mizuuchi, 1986; Craigie and Mizuuchi, 1987; Mizuuchi and Mizuuchi, 1989). The minimal donor DNA sequence which allowed the most efficient catalysis was the pair of Mu R-ends containing the R1 and R2 subsites in inverted orientation (Baker and Mizuuchi, 1992; Craigie and Mizuuchi, 1987; Namgoong et al., 1994). They acted as allosteric effectors and promoted tetramerization of MuA through conformational changes (Namgoong et al., 1994).

Although several normal assembly requirements could be bypassed, the minimal Mu *in vitro* transposition faithfully reproduced the required chemical reactions and the stepwise assembly of the distinct tetrameric transpososome structures with increasing stability to conduct the transposition reaction (Baker and Mizuuchi, 1992; Savilahti et al., 1995). Therefore, Mu transposition reaction can be recapitulated *in vitro* under simple reaction buffer using only a precleaved donor DNA with short 50-bp R-ends or oligonucleotide substrates containing R-ends, MuA transposase and a target DNA as minimal macromolecular components (Haapa et al., 1999b; Savilahti et al., 1995). This minimal *in vitro* reaction has been extensively used for mechanistic studies of transpososome function and organization and used as a tool in various transposon applications (reviewed by Akhverdyan et al., 2011; Chaconas and Harshey, 2002; see below).

1.3.5.1 Structure of the strand transfer complex

The Mu *in vitro* transposition reaction has facilitated crystallization of the long-awaited Mu transpososome structure in an active configuration (Montaňo et al., 2012). Crystallized structure of the final strand transfer complex contains a pair of Mu R-ends oligonucleotides each carrying R1 and R2 binding sites, tetramer of truncated MuA proteins, and a target DNA. The highly intertwined Mu strand transfer complex resembles a pair of scissors, where the Mu end DNAs form the handles and the target DNA the blades (Montaňo et al., 2012; Figure 5 and 7). Similar V-shape for the Mu R-ends and MuA tetramer in type 1 complex has also been proposed by cryo-electron microscopy reconstruction of Mu transpososome (Yuan et al., 2005a). Within the transpososome, MuA monomers bound to R2-ends envelop the inner R1-bound monomers of the same DNA segment, providing a structural role. Mu transpososome crystal structure confirms many earlier predictions and biochemical results such as the complex circuit of macromolecular interactions underlying Mu DNA transposition, catalysis *in trans* with two active sites, and target interaction through IIβ subdomain, but also provides new explanations.

Structure of Mu transpososome contains only the central domains of MuA with the subdomains Iα and IIIβ omitted which are involved in enhancer binding and participate in MuB and ClpX interaction, respectively. However, the transpososome structure resolves a long helix structure for previously uncharacterized IIIα subdomain. Most of the individual MuA subdomains have different functions in protomers bound to R1 versus R2 binding site. In R1 subunits IIIα subdomains pair to form a coiled coil structure, which curves in

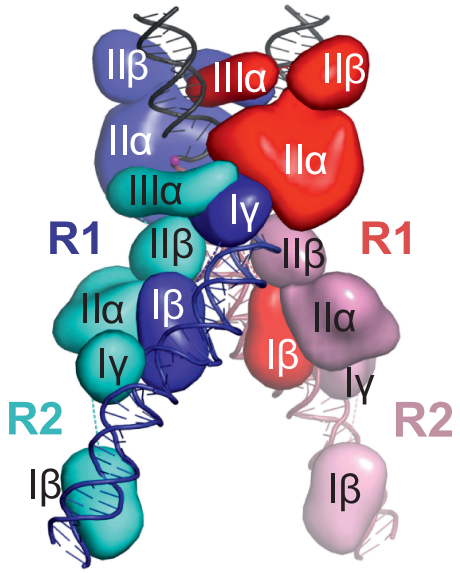

Figure 7 . Mu transpososome structure. Red versus blue subunits are on the opposite sides of the symmetry axis. Catalytic subunits red and blue are R1-bound. Additional subunit pink and cyan are R2-bound. Individual subdomains are shown (Roman numerals and Greek letters). Mu ends are red and blue, and target DNA black. Magenta spheres depict scissile phosphates. The figure is rendered in PyMOL.

the U-shape of the target anchoring it into the complex and stabilizing the strongly bent target. This target bending probably restrains reversibility of the strand transfer and is the basis of the product binding energy. This structural transition of the III domains upon target capture may be also the signal for ClpX remodelling machine. In R2 binding sites, subdomains IIIα of the separate DNA segments wrap horizontally around the other subunits now close to the active site and stabilize transpososome assembly. The Mu transpososome structure proposes that R2-bound IIIα subdomains interact also with the flanking host DNA.

DNA binding domains Iβ and Iγ bind the specific binding sites as confirmed earlier (Namgoong et al., 1998; Zou et al., 1991). However, in the resolved transpososome, they have additional contacts only in R1 subunits, where Iγ has protein-protein contacts with IIIα domain of R2 in the same DNA segments and it contacts IIα domain of R1 subunit on the other DNA segment. However, the modelling the structure provided a depiction for preferential complex assembly also in the natural context with all six DNA-binding sites, enhancer and DNA-bending protein HU involved.

1.3.5.2 Applications of Mu technology

Minimal Mu transposition reaction has several features that make it ideal to be used in a variety of applications. Most importantly Mu can integrate into any target DNA with a high efficiency and a low target site-selectivity (Butterfield et al., 2002; Haapa et al., 1999b; Haapa-Paananen et al., 2002; Mizuuchi and Mizuuchi, 1993; Savilahti et al., 1995). In addition, donor DNA can be custom-designed to include any DNA of variable length between the R-end sequences as well as the R-end sequences themselves can be moderately altered e.g. to contain restriction enzyme cutting sites, translational stop codons or adaptor regions (Brady et al., 2011; Goldhaber-Gordon et al., 2002a; Goldhaber-Gordon et al., 2002b; Haapa et al., 1999b; Jones, 2005; Laurent et al., 2000; Poussu et al., 2004; Poussu et al., 2005; Taira et al., 1999) further extending the application potentiality of Mu transposition system. Many

Table 1. Current applications of Mu *in vitro* system

Application	Use	References
Delivery of primer binding sites	DNA sequencing	Haapa et al., 1999a
	DNA integration site recovery	Brady et al., 2011
Linker insertion mutagenesis	Functional mapping of proteins	Poussu et al., 2004
Deletion and substitution mutagenesis of proteins	Generation of nested sets of N- and C-terminal variants	Poussu et al., 2005
	Deletion of single amino acids	Jones, 2005
	Substitution of single amino acids	Baldwin et al., 2008
	Protein domain insertion	Edwards et al., 2008
	Deletion and insertion of an arbitrary number of bases, scanning mutagenesis and site-saturation mutagenesis	Hoeller et al., 2008
Genetic footprinting	Functional mapping of genome fragments, single genes or proteins	Laurent et al., 2000; Pajunen et al., 2007; Rothenberg et al., 2001; Weber et al., 2010
	Functional mapping of viral genomes	Kekarainen et al., 2002
Genome-wide insertional mutagenesis	Functional mapping of viral genomes	Kiljunen et al., 2005; Krupovic et al., 2006; Vilen et al., 2003
Gene delivery by the use of preassembled transpososome complexes	Genome modification in bacteria, yeast, or mammalian cells	Lamberg et al., 2002; Paatero et al., 2008; P ajunen et al., 2005; Tu Quoc et al., 2007; Wu et al., 2009
Construction of targeting vectors for transgenic animals	Generation of null, potentially hypomorphic, and conditional alleles	Vilen et al., 2001
	Generation of gene knockouts	Zhang et al., 2005
	Generation of gene knock-ins	Jukkola et al., 2005
	Generation of gene knockouts with plasmid-size vectors	Turakainen et al., 2009
Mismatch targeting	SNP discovery	Orsini et al., 2007; Yanagihara and Mizuuchi, 2002

different Mu transposition system applications are based on *in vitro* transposition reaction and introduction of the transposition products into appropriate cells to select integrants and to generate comprehensive libraries of mutant DNA molecules. The current Mu transposition applications include strategies for DNA sequencing, protein engineering, constructing gene-targeting vectors, SNP discovery, genome-wide functional mapping and gene delivery (Table 1).

Specific modification of transposon ends is exploited in designing novel transposon tools. For DNA sequencing application, transposon ends are used as a pair of primer binding sites. Transposon integration events produce primer binding sites randomly throughout the plasmid target and allow DNA sequence to be determined bi-directionally (Haapa et al., 1999a). When transposon ends are engineered to contain adaptor regions complementary to PCR primers and an amine blocking group at the one DNA 3-end, Mu transposition can be used to substitute restriction enzyme cleavage and adaptor ligation (Brady et al., 2011). This method can be used in the field of human gene therapy to recover sites of integrated DNA by DNA barcoding and pyrosequencing (Brady et al., 2011). Similar strategy is applicable to construct DNA sequencing libraries for next generation sequencing platforms.

Mu *in vitro* transposition can be used in engineering of protein-encoding genes in order to study functional domains of proteins and protein-protein interactions. Mu *in vitro* transposition produces random in-frame transposon insertions into a plasmid target containing the cloned gene of study. For linker scanning mutagenesis, Mu transposon ends are modified to contain *Not*I restriction site, which allows the elimination of transposon core sequence from the integration locus and leaving only a 15-bp insert. Within the 15-bp insertion, 10 bp is derived from the ends of the transposon used, and 5 bp from the duplicated target DNA at the insertion site. Translation of the insertions is dependent on the reading frame and the sequence at the insertion site. The analysis of a large number of pentapeptide insertion mutants allows identification of functional or essential protein regions and sites involved in protein-protein interaction as well as insertion tolerant sites (Pajunen et al., 2007; Poussu et al., 2004; Taira et al., 1999; study III). Similar strategy can be used to produce a nested set of N- or C-terminal deletion variants (Poussu et al., 2005). When the same restriction site is contained both in the transposon ends and at the start of the coding region of the target gene, N-terminally deleted fragments are produced after transposon integration and elimination of transposon and the N-terminal part of the target gene by digestion. C-terminally deleted variants can be produced with transposon containing stop codons in three different reading frames at the transposon ends. For deletion of three nucleotides at random positions in a target gene, Mu transposon ends were modified to contain restrictions site for the type IIS restriction enzyme *Mly*I, which allows the elimination of transposon core sequence and 3-bp from the target gene (Jones, 2005). If the removed DNA segment is subsequently replaced with three new nucleotides, the reaction series produces random three nucleotide substitution (Baldwin et al., 2008). Same or similar strategy can be used to introduce a new protein domain (Edwards et al., 2008) and for deletion and insertion of an arbitrary number of bases, which allows scanning mutagenesis and site-saturation mutagenesis (Hoeller et al., 2008) and can be used for cancer gene discovery (reviewed by Copeland and Jenkins, 2010). These protein deletion and substitution strategies are useful in

protein engineering and structure-function studies of proteins (Baldwin et al., 2008; Edwards et al., 2008; Hoeller et al., 2008; Jones, 2005; Poussu et al., 2005).

Mu *in vitro* transposition reaction can be used for functional genetics or genomics of viruses by insertionally mutagenizing either complete or partial genomes cloned in a suitable vector or whole genomes to parallel identify non-essential and essential genes or regions of the genome involved in a biological process (Kekarainen et al., 2002; Kiljunen et al., 2005; Krupovic et al., 2006; Vilen et al., 2003). Genome-wide insertional mutagenesis can be performed also by *in vivo* integration of *in vitro* assembled transpososomes into host genome (Lamberg et al., 2002; Paatero et al., 2008; Pajunen et al., 2005; Tu Quoc et al., 2007; Wu et al., 2009).

Other types of applications that exploit Mu *in vitro* transposition are construction of several types of gene targeting vectors also as a plasmid-size vector to generate transgenic animals (Jukkola et al., 2005; Turakainen et al., 2009; Vilen et al., 2003; Zhang et al., 2005) and to discover single-nucleotide (SNP) polymorphism in DNA by exploiting the preference of Mu to target into mismatched DNA sites (Orsini et al., 2007; Yanagihara and Mizuuchi, 2002). As well, novel applications are continuously in development.

1.4 Random DNA mutagenesis

Random mutagenesis methods are used to introduce genetic diversity throughout a cloned gene of interest in order to create mutant libraries sampling a large fraction of protein sequence. Random DNA mutagenesis followed by selection or screening provides a powerful strategy for protein engineering with an attempt to produce enzymes with improved functions and properties (reviewed Neylon, 2004; Wong et al., 2006; study I–IV). Directed protein evolution takes protein engineering further by imitating natural evolution in a reiterative cycle of genetic variation, phenotypic selection and genotypic amplification aiming to improve enzyme fitness for a given application by altering protein function or evolving a new one (reviewed by Jäckel et al., 2008; Kaur and Sharma, 2006; Kumar and Singh, 2012; Yuan et al., 2005b). The genetic diversity introduced by random mutagenesis is also used to study protein structure-function relationships and functionally analyze uncharacterized open reading frames (reviewed by Neylon, 2004; study IV). Crucial for the success of these studies is the quality of the mutant library most commonly created by random mutagenesis method (reviewed by Dalby, 2011; Shivange et al., 2009; Wong et al., 2006; Wong et al., 2007b; Wu et al., 2009). Traditional random mutagenesis methods can be classified into categories depending on the method employed for nucleotide substitutions: chemistry- or cell-based and enzymatic methods as well as their combinations (Table 2). Classical chemical methods incorporate DNA modifying agents or DNA analogues into DNA and introduce mutations by incorrect replication or repair (reviewed by Griffiths et al., 2000; Miller, 1992; Myers et al., 1985; Singer and Kusmierek, 1982). They are especially efficient when combined with enzymatic methods or used in oligonucleotide synthesis (reviewed by Murakami et al., 2003; Myers et al., 1985; Neylon, 2004; Singer and Kusmierek, 1982; Wu et al., 2009). Relaxing the high fidelity of DNA replication is the basis of mutagenesis also in other methods (reviewed by Neylon, 2004). In cell-based mutagenesis the balance in DNA

replication can be confronted in addition to environmental/physiological stress by defects in one or several DNA repair pathway genes or with a lower fidelity DNA polymerase (Camps et al., 2003; Greener et al., 1996; reviewed by Miller, 1992; Miller, 1996; Miller, 1998). These mutator strains allow easy mutagenesis by *in vivo* propagation of target DNA cloned in replicative vector (reviewed by Neylon, 2004). However, both the chemical mutagens and cell-based mutagenesis accumulate mutations also outside the target gene. These disadvantages are bypassed with enzymatic strategies exploiting DNA polymerase-based error-prone chain reaction (PCR) protocols. Error-prone methods are based on copy errors introduced by DNA polymerases with high error rate and/or reaction conditions imposing them further. The power of error-prone PCR has been behind the success of many directed protein evolution studies. However, it suffers the same challenges as chemical and cell-based methods. All of the methods are biased in terms of mutational spectra, which is assessed by the ratio of transitions (nucleotide substitution that change purine to purine or pyrimidine to pyrimidine) to transversions (nucleotide substitution that change purine to pyrimidine or pyrimidine to purine) (reviewed by Wong et al., 2006; study I). With enzymatic methods there are various efforts to overcome mutational bias such as varying the nucleotide ratio, adding nucleotide analogues, using engineered or natural less biased enzymes (reviewed by Shivange et al., 2009; Wong et al., 2006). However, all of the methods favor transitions over transversion with the outcome of more chemically conservative substitutions instead of chemically diverse substitutions (reviewed by Miyazaki and Arnold, 1999; Neylon, 2004; Shivange et al., 2009).

Ideal circumstances would substitute every amino acid of a proteins sequence by its 19 counterparts in a statistical manner (reviewed by Wong et al., 2006). Although random mutagenesis methods vary in terms of mutational bias, they introduce often only one nucleotide substitution in a codon. Because of the redundancy of the genetic code and organization to minimize mutations, random mutagenesis methods are able to reach only less than 40 % of all the possible changes in protein sequence (Miyazaki and Arnold, 1999; Wong et al., 2006; Wong et al., 2007b). Compared by simultaneous change of two consecutive nucleotides in a codon, it would possible to cover more than 83 % of all possible amino acid changes. In addition, ideal method would be flexible in bias generation e.g. avoiding disruptive amino acid substitutions or stop codons, distribute mutations evenly throughout the gene and have controllable mutation frequency (reviewed by Shivange et al., 2009; Wong et al., 2006; Wong et al., 2007b). All of the techniques in the vast array of random mutagenesis methods vary in terms of nucleotide mutational bias and amino acid substitution pattern, distribution of specific sequences in the library (amplification bias), controllable mutation frequency, dependency on gene length as well as in their technical simplicity and robustness (reviewed by Neylon, 2004; Shivange et al., 2009; Wong et al., 2006; Wong et al., 2006; study I).

Recent advances in random mutagenesis methods overcome the challenge of substituting consecutive nucleotides (reviewed by Shivange et al., 2009). These methods include Sequence Saturation Mutagenesis (Wong et al., 2004; Wong et al., 2008), which employs terminal transferase for incorporation of universal bases and applications based on Mu *in vitro* transposition reaction (chapter 1.3.5.2, trinucleotide exchange (TriNEx):

Baldwin et al., 2008 and transposon integration mediated mutagenesis (TIM): Hoeller et al., 2008). Other powerful strategies to generate molecular diversity include recombination-based methods. Homologous recombination methods recombine genes to form sequence diversity depending on homology and nonhomologous recombination methods without sequence homology (reviewed by Kumar and Singh, 2012). One of the most used techniques is DNA shuffling, which involves random fragmentation of a gene or a pool of homologous genes and recombination in a self-priming PCR reaction (Stemmer, 1994a; Stemmer, 1994b). In addition, to manage challenges in directed protein evolution studies several general theoretical concepts and computational programs are developed to equip protein engineers (Patrick et al., 2003; Volles and Lansbury, 2005; Wong et al., 2006; reviewed by Shivange et al., 2009).

Table 2. Some of the most commonly used random mutagenesis methods resulting in one nucleotide changes

Techniques	Description	References
Chemical method	Chemical methods incorporate DNA modifying agents or DNA analogues into DNA and introduce mutations by incorrect replication or repair.	
Nitrous acid		Myers et al., 1985
Formic acid		Myers et al., 1985
Hydrazine		Myers et al., 1985
Ethyl methane sulfonate, EMS		Lai et al., 2004
Hydroxylamine hydrochloride, NH_2OH-HCl		Stellwagen and Craig, 1997a
Cell-based method	Mutagenesis *in vivo* is performed by transforming a plasmid containing the gene to be mutagenized to a mutator *E. coli* strain. These strains lack DNA repair mechanisms or contain a modified polymerase with lower fidelity.	
E. coli expressing *mutA* allele		Miller, 1992
E. coli XL1–Red strain expressing *mutS*, *mutT*, and *mutD* alleles		Greener et al., 1996
E. coli expressing ep Pol I		Camps et al., 2003

Techniques	Description	References
Error-prone PCR	Genetic diversity introduced by inaccurate duplication of genes by low-fidelity polymerase and/or buffer conditions.	
Taq-pol		Eckert and Kunkel, 1990
Taq-pol unbalanced nucleotides		Leung et al., 1989
Taq-pol unbalanced nucleotides and Mn^{2+}		Cadwell and Joyce, 1992
Taq-pol I614K		Patel et al., 2001
Mutazyme I		Cline et al., 1996
Mutazyme II		Stratagene, 2004
Human X and Y family polymerases		Emond et al., 2008
Pfu-Pol(exo^-)D473G		Biles and Connolly, 2004
Combined methods		
EpPCR followed by 8-hydroxy-dGTP and dPTP by *Taq*-pol		Zaccolo et al., 1996
EpPCR followed by 8-oxo-dGTP mRNA mutagenesis by Qβ replicase		Kopsidas et al., 2007
Ep rolling circle amplification		Fujii et al., 2004

2 AIMS OF THE STUDY

This work focuses on the functional characterization of Mu transposition machinery. Despite the wealth of knowledge of the biochemistry of Mu transposition, little detailed structure-function information is available on the Mu transposition machinery or the determinants that govern the transposition efficiency of the Mu system. The aim of this study was to characterize the MuA transposase protein structure-function relationships and to modify MuA transposase for increased transposition efficiency with the objective of improving Mu *in vitro* transposition applications. For these purposes, it was essential to determine the random DNA mutagenesis conditions suitable for MuA protein engineering and to establish an *in vivo* assay for quantitative analysis of transposition.

The specific aims were:

1. To critically evaluate and compare different random mutagenesis methods. Determine their mutation parameters, controllable mutation frequencies, mutation spectra, and amino acid substitution patterns.
2. To develop a universal and adjustable *in vivo* assay for the quantitative analysis of DNA transposition and to design the assay to mimic the minimal Mu *in vitro* transposition reaction.
3. To quantitatively analyze the structure-function relationships of MuA transposase. Differentiate between the structurally and functionally important regions within the MuA protein that are non-modifiable against those that are modifiable.
4. To generate MuA transposase variants with altered transposition frequencies. Characterize their hyperactivity-inducing properties in the context of the Mu transpososome. Generate MuA variants for different Mu-based applications.

3 MATERIALS AND METHODS

The bacterial strains, plasmids, and oligonucleotides used in this study are described in the original publications. The experimental methods used in this study are described in the original publications and listed in Table 3. References to published methods can be found in the articles.

Table 3. Methods used in this study

Method	Described and used in
Agarose gel electrophoresis techniques	I, II, III, IV
Bacterial transformations	I, II, III, IV
Blue-white screening	I
Cell-based mutagenesis	I
Chemical mutagenesis	I
DNA sequencing and sequence analysis	I, II, III, IV
Electroelution	I, II, III, IV
Error prone PCR	I, IV
FPLC	IV
HPLC	IV
In vitro transposition reaction	III, IV
In vivo transposition assay	II, III, IV
Light microscopy	II, III, IV
Molecular cloning techniques	I, II, III, IV
PCR	I, II, III, IV
Plasmid DNA isolation	I, II, III, IV
Protein expression in bacteria	II, IV
Protein purification	IV
Restriction analysis	II
SDS-PAGE	II, IV
Site-specific mutagenesis	IV
Structural analysis of proteins	III, IV
Western blotting	II

4 RESULTS AND DISCUSSION

4.1 Mutant library generation for protein structure-function studies (I, III, IV)

4.1.1 Random DNA mutagenesis methods (I, IV)

There is a myriad of random mutagenesis methods, which differ in their performance characteristics (reviewed by Jäckel et al., 2008; Kaur and Sharma, 2006; Neylon, 2004; Shivange et al., 2009; Wong et al., 2006; Yuan et al., 2005b; chapter 1.4). As the quality of mutant library is critical to the success of directed protein evolution experiment, an appropriate mutagenic method and specific mutagenic conditions for a particular purpose should be found. Although there are critical reviews and computer programs helping this task, they are based on several independent studies, not on direct comparisons.

For MuA transposase protein engineering purposes and to aid of all protein engineers, we critically evaluated the important performance criteria for the most widely used random mutagenesis methods in the main mutagenesis categories. We varied critical mutagenic conditions and determined relative characteristics for error prone PCR with *Taq* DNA polymerase in low-fidelity buffer condition and with a combination of nucleotide analogues (8-oxo-dGTP and dPTP), error prone PCR with Mutazyme II DNA polymerase using variable amount of amplicon and PCR cycles, XL1-Red mutator strain, and a chemical mutagen NH_2OH-HCl. Our results provide easy to use guidelines for various DNA mutagenesis projects.

4.1.1.1 Visual assay allows convenient means to compare random mutagenesis methods (I)

The properties of the mutagenesis methods were specified using an *E. coli* blue/white screening assay based on the enzymatic activity of β-galactosidase (reviewed by Miller, 1992). This assay allowed a visual and direct comparison of the methods in their operational ranges. We used as mutational target a gene fragment carrying *lacZα*, which encodes α-complementing peptide of β-galactosidase. When transformed into a host carrying N-terminal deletion of *lacZ*, α-peptide complements the defective activity of the host *lacZ* gene product and forms the basis of blue/white screening. When bacteria containing mutagenized reported plasmid library were plated on indicator plates containing X-gal/IPTG, original (or noncritically mutated) plasmids yield deep blue colonies. Plasmid clones harbouring critical mutations with regard to the reporter system, generate white or pale blue colonies. Therefore, the mutation frequency and controllability of a chosen method was determined by the proportion of discernible mutant colonies from the total number of colonies. Although the assay scored only mutant plasmid colonies containing mutations that alter lactose utilization, the proportion of discernible mutant colonies can be considered as a quantitative measure for relative mutation frequencies.

Sequencing was used to analyze the number of mutations along the target sequence in the selected plasmid clones. The number of mutations detected by sequencing correlated well with the relative mutation frequencies. We included *MuA* gene into the mutational

gene target fragment. This allowed specific mutation types and amino acids changes to be determined along a longer region of protein-encoding DNA. In addition, we were able to simultaneously generate random mutations into a *MuA* gene encoding MuA transposase.

4.1.1.2 Error prone PCR allows widely controllable mutation frequencies (I)

The mutation frequencies were determined from homoduplex mutant plasmid DNA libraries by evaluating approximately 1×10^4 colonies according to their colour for each of the mutagenesis method. Error prone PCR protocols proved to be powerful mutagenesis methods with widely controllable mutation frequencies. The relative mutation frequencies observed ranged from 2% to more than 90% when different error-prone PCR mutagenesis methods and various parameters were used (Figure 2 and 4 in I). In the nucleotide level, these specific mutagenic conditions can be used to generate mutant libraries containing 1 to more than 30 mutations within the 1000-bp target gene. *Taq* DNA polymerase with either mutagenic buffer or nucleotide analogues 8-oxodGTP and dPTP was the most efficient mutagenesis method. When the goal is to generate simultaneously many mutations, *Taq* DNA polymerase with nucleotide analogues 8-oxodGTP and dPTP is a desirable protocol to be used. However, in protein engineering studies, the aim is commonly to cause one to four amino acid changes per protein (reviewed by Shivange et al., 2009). For these purposes other error-prone PCR protocols should be used. Even *Taq* DNA polymerase without mutagenic buffer proved to induce sufficient number of mutations for many protein engineering projects. Error-prone PCR with Mutazyme II DNA polymerase allowed generation of mutant libraries of relatively low to moderate mutation frequencies. With the XL1-Red and NH_2OH-HCl protocols the maximum achievable mutation frequencies were far from those obtained with error-prone PCR methods (2.0% and 0.31%, respectively). Regardless of the used mutagenic conditions, these methods are usable for obtaining only a few mutations into a long target gene.

4.1.1.3 Random mutagenesis methods are biased to transitions (I, IV)

All current random DNA mutagenesis methods suffer from a bias towards transition mutations, in spite of the several efforts of compensating their occurrence (reviewed by Shivange et al., 2009; Wong et al., 2006). We determined by DNA sequence analysis the mutation spectrum and the different types of bias at the level of nucleotide mutations for all of the evaluated mutagenesis methods (Table 1 and III, Figure 3 and 5 in I). The number of clones analysed per mutagenesis method ranged from 15 to 27, and the total number of sequenced nucleotides varied between 12,000 and 21,000 per method.

Transition bias was prevalent among the methods analyzed. *Taq* DNA polymerase with nucleotide analogues and NH_2OH-HCl protocols were the most biased methods and Mutazyme II DNA polymerase the least biased. With regard to bias between different transition mutations, *Taq* DNA polymerase was biased to A:T→G:C and all the other methods to G:C→A:T transitions. The bias among transversion mutation was commonly towards A:T→T:A transversions. Mutazyme II DNA polymerase was the only method able to induce all types of mutations, also unwanted 1-bp insertions and deletions. Deletions of 1 bp were detected also with *Taq* DNA polymerase protocols and XL1-Red mutator strain. The sequence analysis in study IV for MuA mutants generated by *Taq* DNA polymerase

with mutagenic buffer protocol and Mutazyme II DNA polymerase with varying number of PCR cycles allowed supplementary validation and confirmation of the mutational spectrum analysed (data not shown).

4.1.1.4 Amino acid substitution patterns (I, IV)

Nucleotide mutations induced by a given mutagenesis methods influence the changes detected at the protein level. We evaluated the amino acid substitution patterns generated by the different error-prone PCR methods. We used compiled mutation data from the protein-encoding regions of the LacZα sequence (60 amino acids) and C-terminus of the MuA sequence (143 amino acids) (Table 2 in I). The total number of amino acids analyzed varied among the different methods from approximately 3000 to over 4200. Mutations were distributed evenly along the target genes. *Taq* DNA polymerase with nucleotide analogues 8-oxodGTP and dPTP was the only method to produce consecutive nucleotide mutations, because of the high mutation frequency. Mutations were classified into those that preserved the corresponding amino acid (silent mutation), changed it (missense mutation), or generate a stop codon (nonsense mutation). The differing preferences towards transitions versus transversions with *Taq* DNA polymerase and Mutazyme II DNA polymerase based methods caused differences also at the protein level. *Taq* DNA polymerase methods, which are more biased to transitions, preserved the amino acid more often than Mutazyme II DNA polymerase methods. This was influenced by the organization of genetic code, as it is optimized in the way that transitions lead often to chemically similar or identical amino acids (reviewed by Shivange et al., 2009; Wong et al., 2007a). Therefore, when produced amino acid changes were divided further to conservative and nonconservative substitutions, transition bias influence was obtained again. Mutazyme II DNA polymerase generated more detected nonconservative substitutions. However, when amino acid substitutions were analysed according to the change of chemical group, the different mutagenesis methods induced a very similar substitution pattern. This was detected also for the sequenced hyperactive MuA variants generated by *Taq* DNA polymerase and Mutazyme II DNA polymerase in study IV (Figure 3 and Supplementary Table 5 in IV).

It is considered that for successful protein engineering, it is important to avoid causing structurally interfering changes or stop codons (reviewed by Shivange et al., 2009; Wong et al., 2006). We analyzed the proportion of destabilizing amino acids substitutions such as glycine and proline as well as the stop codons. The total amount of these changes was similar with *Taq* DNA polymerase and Mutazyme II DNA polymerase. Because Mutazyme II DNA polymerase based methods yielded biases towards A and T nucleotides, they had a higher probability of generating stop codons. Correspondingly, *Taq* DNA polymerase methods were biased towards G and C nucleotides and generated more Gly codons and Pro codons. These structurally interfering glycine and proline changes were detected also for the screened MuA variants in study IV (Figure 3 and Supplementary Table 5 in IV). Generated proline residues were located in protein region not important for activity in the assay used. But many glycine residues generated improved activity. Therefore, too rational protein engineering may miss unexpected outcomes. One stop codon detected was proven to be leaking and allowed full length enzyme production (data not shown).

4.1.2 Random MuA mutant plasmid libraries (IV)

The genetic variation generated in study I allowed us to choose gene libraries containing variable number and types of mutations within the *MuA* gene. A total of five different MuA mutant libraries were generated using *Taq* DNA polymerase under three mutagenic PCR conditions with 0, 1 or 2 µl of mutagenic buffer and Mutazyme II DNA polymerase employing 5 and 10 cycles of amplification. The resultant five mutant libraries included a total of ~3×10^5 independent plasmid clones (Supplementary Table 4 in IV). Of these libraries, 64,000 colonies were screened for enhanced activity (chapter 4.2.4.3). The amount of colonies exhibiting hyperactive phenotype ranged from 1.0% to 3.4% among protocols. Independent random mutant libraries generated with *Taq* DNA polymerase produced more hyperactive variants than with Mutazyme II DNA polymerase. In addition, the two clearly most hyperactive variants were from *Taq* DNA polymerase mutant library (Supplementary Table 5 in IV). Overall, both Mutazyme II DNA polymerase and *Taq* DNA polymerase protocols generated the same important amino acid substitutions in many critical positions. Commonly, transversions can introduce more variety into the mutant library. However, the compensation of transition bias with Mutazyme II DNA polymerase did not seem to affect MuA protein engineering results. These data indicated that the quality of random mutant libraries were high for protein engineering purposes. At the time we began this study, methods that introduce randomly consecutive nucleotide mutations in a codon were not in common use. Today, for example Mu *in vitro* transposition can be used to randomly replace amino acid residue to all possible counterparts (Baldwin et al., 2008). However, these kind of changes are commonly required if the intention is to derive new activities for a protein. If the aim is to improve the protein like in this study, subtle changes (caused by transitions) in protein sequence generally accomplish the task.

4.1.3 MuA insertion mutant plasmid library (III)

Functional regions in proteins can be studied via transposition-assisted generation of libraries of linker insertion mutants (chapter 1.3.5.2). We decided to study MuA structure-function relationships using pentapeptide scanning mutagenesis (Poussu et al., 2004). We used Mu *in vitro* transposition reaction to introduce five amino acid insertions randomly throughout the MuA protein sequence. Initially, we generated a pool of plasmids each containing a transposon insertion within the *MuA* gene. Subsequently, we eliminated the transposon core sequence by digestion and recircularized the plasmid backbone to generate a pool of plasmids each containing only a 15-bp insertion within the *MuA* gene. The resultant MuA mutant plasmid library included a total of ~6×10^4 plasmid clones. The isolated plasmid DNA from 608 clones was subjected to restriction analysis to roughly map the 15-bp insertion sites. On the basis of the initial screen, a total of 331 clones were selected for insertion site mapping by DNA sequence analysis. From the large set of sequenced clones, a total of 233 clones with unique insertion site were identified (Supplementary Table 1 in III). The overall distribution of insertion sites and their localization within the secondary structural elements allowed a comprehensive structure-function analysis of MuA (Figure 4 in III).

4.2 Papillation assay –means to screen mutants (II, III, IV)

The success of protein engineering projects relies on the quality of mutant libraries and screening and selection strategies (reviewed by Kumar and Singh, 2012). Screening and selection methods should allow robust and high-throughput detection of even the comparatively slightly improved protein variants from the large mutant libraries. In order to monitor MuA mutants with altered transposition frequency, we needed a convenient assay to compare activities directly as well as quantitatively.

There are several transposition assays developed in the past for *in vivo* analysis of transposition frequency (reviewed by Miller, 1992; Snyder and Champness, 2007). The most widely used method to study *in vivo* transposition is a visual colony screening assay also suitable for large-scale studies. This assay exploits the formation of coloured micro colonies or papillae within otherwise colourless bacterial colonies (Huisman and Kleckner, 1987; Stellwagen and Craig, 1997b). Such papillation assays employ *lac* operon and utilize a reporter transposon including a promoterless *lacZ* gene (Derbyshire and Grindley, 1996; Huisman and Kleckner, 1987; Krebs and Reznikoff, 1988; Lee and Harshey, 2001; Stellwagen and Craig, 1997b). Integration event into a genomic locus under the control of an active promoter would generate an active gene fusion and result in the expression of the *lacZ* gene. These events are detectable by the appearance of coloured Lac^+ papillae on colourless Lac^- colonies on appropriate indicator agar (Reznikoff et al., 1993). The number of papillae per colony is proportional to the frequency of transposition events catalysed by the transposase. This strategy has been used to study several host- (Makris et al., 1988; Swingle et al., 2004; Twiss et al., 2005; Weinreich et al., 1994), and element-encoded (Bender and Kleckner, 1992; Derbyshire and Grindley, 1996; Kim and Harshey, 1995; Lu and Craig, 2000; Naumann and Reznikoff, 2002; Weinreich et al., 1994; Wiegand and Reznikoff, 1992) functions and transposon end sequences (Huisman and Kleckner, 1987; Lee and Harshey, 2001; Makris et al., 1988; Tang et al., 1995; Tavakoli and Derbyshire, 1999) for a variety of transposons. However, all current papillation assays are element specific and do not allow quantitative or a broad dynamic range detection of transposition events.

We decided to exploit papillation for development of an easy-to-use method, which allows quantitative analysis of transposition frequency and is applicable to a variety of mobile DNA elements. These types of transposition assays are needed to characterize newly detected mobile elements as well as for methodology development of different transposons.

4.2.1 Characteristics of the papillation assay plasmid (II)

Our aim was to set up a universal platform that is based on the transformation of *E. coli* with a single plasmid containing all the components required for transposon mobilization. First, we constructed the papillation assay plasmid, pLHH4, for phage Mu (Figure 1A in II). Important features of the papillation assay plasmid include (i) *MuA* transposase gene under arabinose/glucose-controllable P_{BAD} promoter of *E. coli* and the expression unit bordered by strong transcription terminators, (ii) mini-Mu transposon containing N-terminally truncated *E. coli lacZ* ($\Delta_{1\text{-}8}$*lacZ*) and *cat* gene between the 50-bp segments of Mu R-end DNA. Second, we modified the papillation assay plasmid pLHH4, to be used as a universal

papillation plasmid cloning vector. In this modified papillation assay plasmid, pSKT1 (Supplementary Figure 2 in II), Mu specific sequences were replaced by polylinkers. Therefore, straightforward directional cloning of any mobile element ends and a cognate transposase gene into the polylinkers enable adjustable *in vivo* transposition analysis of any specific DNA transposon system. These pBR322-derived plasmids exhibit medium copy number, and encode β-lactamase. Two antibiotic selection marker genes prevent the growth of satellite colonies on plates emerging during long incubation period. In addition, marker gene within the reporter transposon may be useful feature in downstream applications.

4.2.2 Performance features of the papillation assay (II, IV)

The papillation assay is based on visual scoring of the transposition events. After transformation of the papillation plasmid into Lac$^-$ *E. coli* strain, several events lead to the quantitative analysis of transposition: (i) The expression of transposase protein results in the initiation of the transposition reaction. Transposase binds to the transposon ends and mobilizes the reporter transposon. (ii) When transposon is integrated into an expressed chromosomal gene in a correct orientation and reading frame it results in *lacZ* gene fusions and an expression of β-galactosidase fusion protein. (iii) These events are detectable on the colour indicator plates after prolonged incubation by the formation of blue papillae on otherwise whitish colonies. A single papilla represents an individual transposition event. Our aim was to create assay conditions for a reliable and discernible papillae count. We also aimed to adjust the transposition frequency to a desired level via the control of arabinose promoter in the papillation plasmid. Standard papillation plates contained arabinose, X-gal, IPTG, and lactose.

4.2.2.1 Papillation characteristics

We evaluated different phenotypically Lac$^-$ *E. coli* strains including DH10B, DH5α, and JM109 for their suitability as a standard host in papillation assay (Table 1 in II). In addition, HT321 stain was evaluated as a low-papillation host. For the analysis, plasmid pLHH4 was transformed into various Lac$^-$ *E. coli* strains and colonies were grown on indicator plates with varying arabinose concentration for a reference time period. A number of critical attributes affecting the assay performance were evaluated, which proved to be strain-dependent. DH5α strain formed large colonies with even distribution of papillae within a colony. Colonies of DH5α strain allowed scoring of more than 500 discernible papillae. In addition, the papillae number in DH5α strain was adjustable over a wide range of arabinose concentrations and was not significantly affected by the total colony number on a plate. With regard to these aspects, DH5α strain was considered optimal for *in vivo* papillation assay.

For further optimization and to verify the maximum dynamic range, plasmid pLHH4 was transformed into DH5α strain and colonies were grown on indicator plates containing arabinose concentration of 0.1% at different temperatures: 22ºC (room temperature), 25ºC, 30ºC, and 37ºC. Emerging papillae were enumerated as a function of incubation time (Figure 2 in II). At two higher temperatures large dynamic range was observed, but two lower temperatures were unable to produce reasonable maximum transposition. At the temperatures 30ºC and 37ºC, the papillae number increased linearly between the certain

time points. As incubation at 30ºC allowed more control to papillae formation, it was used in further studies as standard incubation temperature as well as 115 h as standard incubation time. However, lower temperatures may be required for stringent condition analysis of highly active transposase variants (Figure 7B and Supplementary Table 5 in IV).

4.2.2.2 Adjustability

The expression level of MuA transposase should impact the transposition frequency which is measurable by the number of papillae per colony. Initially, we examined the MuA expression levels under variable arabinose concentrations in liquid cultures. Subsequently, transposition frequency was studied under variable arabinose concentrations in papillation assay. The population-average protein expression levels of MuA analyzed by SDS-PAGE and Western blotting indicated that MuA expression levels were directly proportional to arabinose concentration up to 1×10^{-2}% (Figure 3A in II). In papillation assay, papillae begun to be visible at ~1×10^{-4}% arabinose concentration and continued to form up to 1×10^{-1}% arabinose concentration (Figure 3B in II). These results demonstrate that MuA expression level is induced by the arabinose concentration and correlates with a papillae number over a wide range of arabinose concentrations. It is known that at subsaturating arabinose concentrations gene expression from plasmids containing P_{BAD} promoter result in mixed population of cells that are either highly induced or uninduced (Siegele and Hu, 1997). Yet, a degree of transposase expression variation is present in the subpopulations, resulting in a controllable population level average expression of transposase by arabinose concentration. The adjustability of the transposition frequency allows assay to be used for different screening purposes. For example, for the screen of more active mutants, the wild type activity level is adjusted to yield only a few papillae per colony (IV).

Glucose can be used to add another level of control to transposase expression. P_{BAD} promoter was observed to allow some basal expression of transposase. If uninduced transposase expression results in a high papillae number, glucose can be added onto growth medium to reduce the papillae number to a desired level (see below).

4.2.3 Universal platform (II)

The aim was also to develop the assay to applicable for other mobile DNA elements. We verified the universality of the assay by cloning of the critical components of IS*903* transposition system into the universal papillation assay cloning vector pSKT1. This papillation assay plasmid of IS*903* was transformed into DH5α cells and analysed for papillation in standard assay conditions under varying concentrations of arabinose. High transposition frequency was observed for the natural IS*903* system, even without arabinose induction. The addition of glucose at varying concentration allowed transposition frequency to be adjusted linearly (Figure 6 in II). Therefore, the papillation assay established could be used for *in vivo* transposition analysis also for other transposition system. Although Mu and IS*903* both use replicative transposition, we presume this papillation assay to be used for other types of transposable elements as well.

4.2.4 MuA mutant screening in papillation assay (II, III, IV)

The primary aim in the papillation assay development was to provide means to screen and scrutinize different types of MuA transposase variants, also potentially usable for Mu *in vitro* transposition applications. Therefore, the assay was designed to mimic minimal *in vitro* reaction as closely as possible. Accordingly, reporter transposon contained only 50-bp of Mu R-end in each transposon end. The system omitted the known auxiliary factors of Mu transposition, such as the transposition enhancer and activator/targeting protein MuB. The assay scores only fully productive transposition events and allow reliable assessment of determinants affecting transposition.

4.2.4.1 Different MuA activities can be quantitatively analyzed in papillation assay (II)

Our aim was to test the assay under different levels of transposase activity and thus verify applicability of the assay as a screening method in protein engineering studies. We examined in the papillation assay various deletion variants of MuA previously scrutinized *in vitro* conditions (Baker et al., 1991; Clubb et al., 1996; Leung and Harshey, 1991). We used an N-terminally deleted $MuA_{77\text{-}663}$ and C-terminally deleted variant $MuA_{1\text{-}615}$ as well as a variant missing both termini $MuA_{77\text{-}615}$. In addition, DDE-triad mutant MuA_{E392Q}, which is proficient for the transpososome assembly but catalytically defective, was used (Baker and Luo, 1994). MuA variants were analyzed under varying arabinose concentrations with standard papillation condition 30ºC, 115 h (Figure 4 and Figure 5 in II). The assay allowed activities above and below the wild type level to be reliably detected. N-terminal variant was hyperactive, DDE-triad mutant inactive, and C-terminal deletion containing proteins were hypoactive. The papillation assay lacks the auxiliary transposition factors enhancer sequence and MuB protein, which are bound by subdomains Iα and IIIβ, respectively. When the same transposition factors are omitted *in vitro*, subdomains Iα and IIIβ are not essential for the transposition reaction (Kim and Morrison, 2009). In the enhancer independent conditions subdomain Iα have even been shown to be somewhat inhibitory (Mizuuchi and Mizuuchi, 1989; Yang et al., 1995) and under these conditions deletion of the subdomain Iα has induced hyperactivity (Clubb et al., 1996; Kim and Morrison, 2009). However, the subdomain IIIβ is also bound by ClpX, which is required for transpososome disassembly. Inability of this function conceivably causes hypoactivity *in vivo*. Overall, results demonstrated that Mu transposition can be reproduced with minimal components also *in vivo*. Most importantly, the assay proved to be sensitive for advanced screening purposes.

4.2.4.2 Broad range of activities among insertion variants of MuA (III)

MuA insertion variants generated by pentapeptide scanning mutagenesis were analyzed in standard papillation conditions at 30ºC for 115 h with arabinose concentration of 0.1%. Under these conditions, wild type MuA produces ~200 papillae in a colony. Therefore activities above and below a wild type level are easily detected. Insertion variants exhibited a broad range of activities in papillation assay (Figure 3 and Figure 4 in III). Most of the insertions reduced the MuA activity and even more than half of the insertion variants (125 of 233) were totally inactive with no papillae produced. However, many mutants (36 of 233)

had an activity close to wild type (activity level 70 -130 %) and some insertions variants (12 of 233) had activities clearly above the wild type level. Even two-fold activity over the wild type level was observed. Results indicated that in general pentapeptide insertions significantly interfere with the protein structure and/or function. However, certain regions in MuA protein allow insertional modification.

4.2.4.3 Random mutations in MuA can induce hyperactivity (IV)

Different random MuA mutant libraries were screened for hyperactive MuA variants in papillation assay conditions at 30ºC for 115 h with arabinose concentration of 1×10^{-4}%. Under these conditions, wild type MuA produces ~10 papillae in a colony and allows straightforward detection of different types of hyperactive variants. A total of 64,000 colonies were screened by visual inspection. A large set of clones portraying a high activity phenotype were selected for further scrutiny. The 89 plasmid clones, which clearly induced a substantially elevated *MuA* specific level of papillation, were subjected to quantitative papillation analysis under stringent conditions (at 25 ºC for 140 h). In addition, their *MuA* sequence was determined and translated to reveal changes at the amino acid level. A total of 72 unique sequences were identified, and most of these sequences contained several changes (Supplementary Table 5 in IV). The variants portrayed a broad spectrum of hyperactivity, and a correlation with specific amino acid changes was apparent (Figure 3 and Supplementary Table 5 in IV). Certain mutations were identified frequently and special hot spots could be discerned. As these mutations were generated independently in different libraries their role in causing hyperactive phenotype was proposed. In contrast, many changes that were identified only once were supposedly irrelevant for the hyperactive phenotype. The results suggested that the screen used was sensitive and effective in revealing hyperactivity-causing mutations. To dissect the potential mutations causing hyperactivity phenotype, we generated single-substitution MuA variants. We included changes that were present in several independent clones or present in an individual clone yielding extremely improved activity. In addition, we wanted to study mutations located in all of the separate subdomains. A total of 47 single substitutions in 35 different amino acids were generated and their the transposition activity was quantified in papillation assay (at 30 ºC for 115 h with arabinose concentration of 1×10^{-4}%) (Figure 4 in IV). We identified 34 substitutions in 27 specific amino acid residues, which enhanced the protein activity at least two-fold. The highest activity detected was 50-fold over the wild type level. This indicated that we were able to find the particular residues and specific changes in them that resulted in hyperactivity from the large set of mutants. Although mutant libraries produced by error-prone PCR cannot contain an exhaustive mutant repertoire, every residue in the protein was subjected to a change. We suppose that most of the amino acids in MuA protein sequence potential to cause hyperactivity phenotype were detected.

The highest papillation activities were detected for screened variants containing multiple substitutions. Therefore, it was evident that several activity-improving substitutions can be combined to yield synergistic effects on the transpositional activity of MuA. To scrutinize synergism, various mutation combinations were generated with the important amino acids contained in the most active variant identified in our screen (clone EP3I4)

(Figure 7, Supplementary Table 5 and 6 in IV). Of the five amino acid substitutions, three improved activity as single-substitution variant and were combined. Triple-substitution mutant (W160R, W345R, M374V) portrayed the highest activity reaching that of the original variant. To further validate the effect of various mutation combinations M374V was changed to E233K. This variant containing mutations in three different subdomains was the most active variant and produced over 500 times more papillae than wild type MuA (Supplementary Table 6 in IV). The data suggest that finding a special combinations of different mutations yield high activities due to synergism, an effect that has also previously been reported in other transposase protein engineering studies (Goryshin and Reznikoff, 1998; Lampe et al., 1999; Mates et al., 2009; Yusa et al., 2011; Zayed et al., 2004).

4.3 MuA structure-function relationships (III, IV)

We used two different approaches to study MuA protein structure-function relationships and to generate more efficient MuA transposase variants. Both of the methods used for mutant library construction, random mutagenesis and pentapeptide scanning mutagenesis, are straightforward strategies for library construction without the requirement for structural information of the protein. At the time we started this work, detailed structural information was available only for the separate subdomains of domain I and II. However, the recently resolved Mu transpososome structure with Mu R-ends and including target DNA allowed us detailed architectural probing of the functional data acquired by *in vivo* papillation analysis. The combined structure-function analysis forms a platform for further rational protein engineering.

4.3.1 Insertion tolerant versus intolerant regions within MuA domains (III)

The correlation of the activities of the 233 MuA insertion variants with the MuA domain structure (Figures 3–5 in III) produced a comprehensive map of insertion tolerant versus insertion-intolerant regions in the MuA protein. We observed clear differences in insertion tolerance between the separate subdomains of MuA (Figure 3 in III). The terminal subdomains Iα and IIIβ of MuA allowed insertions well. In contrast, subdomains Iγ and IIIα appeared to be entirely insertion-intolerant. Other subdomain Iβ, IIα, and IIβ tolerated only some insertions in certain confined regions. A high degree of malleability (wild type level of activity or more) was detected for inter-domain linker connecting Iα to Iβ and in certain loops (Iβ: aa 126–130; IIα: aa 280 and aa 474–479) connecting secondary structural elements (Figure 5 and Figure 6 in III). Also several other loop regions allowing insertions with reduced activity were detected. For further validation of the quantitative functional data, we produced a protein sequence alignment for 44 members of MuA family transposases. In essence, the two separate data sets revealed the same modifiable regions in the MuA family members (Figure 6 and Figure 7 in III).

4.3.1.1 Effects of insertions in the functional core of Mu transpososome

Our analysis revealed that the structure of MuA does not easily allow insertional modification. Apart from terminal subdomains of MuA, the only allowed insertions were in loops or at the ends of secondary structural elements. This was consistent with the general scenario of insertion accommodation: Insertions are commonly accommodated on protein surfaces and in mobile protein regions. Inside of the protein core, amino acid chain extension is tolerated if it does not severely distort the protein structure (Heinz et al., 1994). Function of MuA is especially susceptible for structural rearrangements as it functions as a tetramer in the context of a large protein-DNA complex. There are several functions assigned for the MuA in the transpososome. In addition, protomers of MuA tetramer have separate roles when bound to R1 sites versus R2 sites of the transposon ends. The insertion variants were analysed in the papillation assay, which measures the outcome of the transposition. Therefore, the insertions had to be tolerated in each of the protomers within the tetramer.

The insertion tolerance of subdomains Iα and IIIβ was consistent with the results derived with N- and C-terminal deletion variants in the papillation assay (chapter 4.2.4.1). Most of the indels found in the sequence alignment data of protein homologues were also in these regions. Results suggest some degree of malleability in their functions. Because subdomain Iα is a DNA-binding module, distortion of this activity under conditions where the specific binding is not possible may improve the activity via inhibition of nonspecific DNA binding events. Correspondingly, insertions to the linker between Iα and IIβ may allow subdomain Iα to be moved further away causing less interference.

In contrast to enhancer binding, the recognition and specific DNA binding to transposon ends via subdomains Iβ and Iγ is fundamental for the activity. Accordingly, stringency in the architectures of these domains was detected (Figure 5 in III). Subdomain Iβ did not tolerate insertions within its DNA-binding surface or in the structures anchoring the recognition helices and subdomain Iγ was entirely insertion intolerant. In addition, insertions into the linker region between the domains suggested to bind to the DNA minor groove almost totally abolished the activity. High sequence conservation has been shown for these domains (Montaňo et al., 2012). Same conservation was revealed by our combined data from the structure-function analysis and protein sequence alignment. Because the subdomain Iγ has also several other important protein-protein and protein-DNA contacts in transpososome when bound to R1 subunit, any structural perturbation to this helix bundle has to be functionally compromised.

High conservation for the RNase H-like fold and the catalytic core was detected (Figure 5 and Figure 7 in III). However, insertions into the loop residue aa 282 connecting first and second β-strands of the RNase fold improved the activity. As this loop interacts with subdomain IIβ, the enhanced activity is supposed to increase the flexibility between the subdomains of the catalytic core. Another loop region (aa 474–479) allowing insertions with improved activity is in R2-bound subunit involved in protein-protein interaction with Iβ domain of R1-bound subunit. The only allowed insertions to subdomain IIβ were in one loop tip (aa 529–531), which is not involved in target DNA capture. Mu transpososome structure revealed two separate roles for the long helix structure of subdomain IIIα when bound to R1 versus R2 subunit. These functions were prevented by insertions into the helix, indicating

their essentiality for transposition. Generally, we propose that the allowed insertions in the inner MuA domains (Iβ to IIIα) were involved in the formation of favourable protein-protein contacts within the transpososome.

4.3.2 Hyperactivity-inducing substitutions in MuA (IV)

The MuA mutant screen identified 34 substitutions in 27 specific amino acid residues, which enhanced the protein activity at least two-fold (Figure 4 in IV). Although, most of these substitutions were in catalytic subdomain IIα, also all other subdomains contained at least one substitution enhancing MuA activity (Figure 5 in IV). The highest activities were detected for substitutions located in subdomains Iγ, IIα, and IIβ. The only enhancing substitution located in subdomain Iα was A59V. It increased the protein activity to some extent (2.8-fold). As discussed previously (chapter 4.2.4.1) hyperactivity caused by mutations in Iα supposedly weaken the disturbing nonspecific DNA binding in enhancer independent conditions. In subdomain Iβ, two highly activating substitutions D97G and W160R are located in different helix structures not involved in DNA binding. The most active substitution and abundant in all mutant libraries was located in subdomain Iγ. There were two different substitutions for this residue with different electrostatic changes, E233K and E233V. Other detected substitutions (E179A, E179V, D232N) in subdomain Iγ increased protein activity only slightly. Residues E233 and D232 are located in the turn/loop region following DNA recognition helix but not close to DNA. Residue E179 is located in the DNA minor-groove-binding linker region between Iβ and Iγ.

The most of the activating substitutions located in subdomain IIα enhanced the activity more than 15-fold (Q254R, I335T, G340S, W345C, M374V, F447S, R478C, R478H, E482K, E483G, E483V, M487I). Also substantial enhancement was detected for several substitutions: E258G, G302D, W345R, M374T, F447Y, and F464Y. Two substitutions with only a slight activity increase were D320V and D466G. In domain structure of IIα, these residues can be divided to two regions, the one is the several loops and secondary structures surrounding DDE motif and the other is the loops and secondary structures on the surface on the other side of the domain. In subdomain IIβ, two substitutions V495A and V507A enhanced activity extensively (14-fold and 26-fold, respectively) and two substitutions Q539H and Q539R moderately (5-fold and 6-fold, respectively). V495A and V507A and are located in β-strands near the turn from IIα that forms part of IIβ. Residue Q539 is located in a helix also close to subdomain IIα.

One activating mutation was located in each of the separate subdomains of domain III. Substitution Q594R had 10-fold activity and I617T only 2.5-fold activity. Substitution Q594R is located at the end of a long helix structure of subdomain IIIα involved in DNA binding. The structure of subdomain IIIβ has not been resolved and was not involved in the Mu transpososome crystal structure. Because, substitution of I617T is in the region participating ClpX binding (Abdelhakim et al., 2008), it is supposed to be beneficial at protein-protein interactions.

4.3.2.1 Hyperactivity characteristics in Mu transpososome

Mapping of the hyperactivity-inducing substitutions into the Mu transpososome structure, offered us insights into mechanisms of hyperactivity. We were able to reveal several different protein-DNA and protein-protein interaction surfaces where mutations resulted in hyperactivity. In the Mu transpososome structure, hyperactivity-inducing mutations were found to locate in two different interface groups between R1-bound and R2-bound subunits. Many mutated residues were within the interaction surface between subdomains IIα and Iβ (aa 97, 160, 478, 482, 483 and 487) (Figure 6C in IV). Substitution of residues to less negative or more positive in this interface may conceivably improve protein-protein interactions. Because subdomain IIα has a structural role in this context, mutations may offer better stability for the structure. This region was found also in study III to result in hyperactivity via insertions. Another R1–R2 interaction group was detected for IIβ and IIIα subdomains of R2 subunit sandwiching the turn/loop region following the recognition helix of Iγ domain (aa 232 and 233) of R1 (Figure 6C in IV). This interface contained the two most active substitutions of the same residue (E233K and E233V), indicating importance of the charge change in the protein-protein interface. This interaction has been proposed to be important as there is high sequence conservation for subdomain Iγ in MuA related sequences (Montaňo et al., 2012; study III).

The hyperactivity-inducing substitutions in subdomain IIα were close to the catalytic core (aa 302, 335, 340, 345, and 374) or packed against loops containing E393 and/or the adjacent loop interacting with non-transferred strand (aa 254, 258, 447, 464, and 466) (Figure 6D in IV). Because several of the substitutions close to active site changed the packing and flexibility properties of the affected residue the proposed effect was involved in packing of the core. Similarly, the loop containing E392 must change conformation for catalysis. Therefore, it is suggested that mutations close to that loop and/or the adjacent loop must be involved in conformational changes.

Improved DNA binding is assumed for two substitutions changing the electrostatic of the residue to more desirable for DNA binding. E179V in subdomain Iγ is close to Mu end DNA on R1 and R2, and close to the proposed flanking host DNA on R1. Correspondingly, substitution Q594R in subdomain IIIα is close to target DNA on R1 and the proposed flanking host DNA on R2.

Although substitutions in MuA structure can effect on general properties of the protein, improvement of the protein activity must also reflect critical functional aspects. Particularly, multiple substitutions were localized to the same hyperactivity-inducing surfaces as well as were complemented by the insertional mutant data and the evolutionary component offered by sequence alignment of protein homologues. Therefore, we propose the activity enhancements are explained as being associated with DNA binding interactions, active site packing and conformational changes, as well as subunit-subunit interface associations. All these interactions have an effect in higher-order transactions, from assembly to structural transitions during the progression of transpososome development proficient for catalysis.

4.3.3 MuA substitution mutants for application purposes

Hyperactivity-inducing substitutions and their activities were identified using the *in vivo* papillation assay. Despite the assay was designed to mimic Mu *in vitro* transposition, it reflects transpositional activity *in vivo*. For *in vitro* activity assessment, 30 single-substitution MuA protein variants and a triple mutant (W160R, E233K, W345R) were purified and evaluated in Mu *in vitro* application settings (Figures 8 and 9, Supplementary Figures 2 and 3 in IV). First, Mu *in vitro* transposition reaction was performed for each of the protein variants with precleaved mini-Mu transposon and pUC19 plasmid as a target DNA (Figure 8A in IV). Second, Mu transpososomes were assembled *in vitro* in the absence of divalent cations. Following electroporation into *E. coli* cells, these complexes integrate transposon DNA into the host chromosomal DNA (Figure 9A in IV). Reaction products were analyzed qualitatively (Supplementary Figure 2 and Supplementary Figure3 in IV) by agarose gel electrophoresis and scored quantitatively by antibiotic selection on plates (Figure 8B and Figure 9B in IV).

Most of the activities in both assays were above the wild type level but the scale of activities was narrower than detected *in vivo*. The most active single-substitution variant had activities up to 6-fold and 4-fold in two different assays. Similar activity reductions for *in vivo* screened hyperactive transposases have been detected *in vitro* also in other transposition systems (Lampe et al., 1999). The first assay resembles *in vivo* papillation assay as both emulate the entire transposition pathway. In this assay the activities of the purified proteins mostly compared well to the relative activities detected *in vivo*. However, five variants harboring a mutation within a stretch of amino acids encompassing residues 320-345 had a substantial reduction of activity compared to the activities detected *in vivo*. In the second assay, the activities of the purified proteins differed more from their *in vivo* activites. In this assay, also the execution of the transposition pathway differs more from the *in vivo* papillation assay as the transposition is phased to seperate complex assembly and subsequent integration event. However, the amounts of complexes produced by various protein variants were consistent with the number of integration events. All the proteins except D320V were capable for transpososome assembly and intergration. Approximately two thirds of the proteins resulted in enhancements in genomic integration. The triple mutant (W160R, E233K, W345R), which portrayed hyperactivity *in vivo* due synergism generated synergistic activity also *in vitro*.

Some of the differences between results of *in vivo* and *in vitro* assays can be explained by the oxidizing *in vitro* versus reducing *in vivo* conditions. Especially, substituted W345C may get oxidized under the *in vitro* assay conditions causing interference. The differences between results may also reflect host factor involvement present *in vivo* but not *in vitro* conditions. As both assays were conducted with one protein preparation per each substitution, the quality of the preparation may affect in principle. We conducted protein gel analysis and nuclease assay as quality control measures, but other protein feature measurement should be performed to validate the results. However, as most of the protein variants behaved in the assays very similarly, at least for the majority of the proteins, the results should be very reliable.

Although, activity levels detected for purified proteins *in vitro* were lower than *in vivo* conditions, the 6-fold and 4-fold activities detected in two different *in vitro* assays are significant in many application purposes. Results indicate that to encounter the highest activity for separate *in vitro* applications, an application specific blend of mutations should be introduced in view to include all hyperactivity interfaces revealed in this study. In addition, to combine hyperactivity data from other studies such as deletion of N-terminal domain and including hyperactive insertions could produce a new level of activity improvement also *in vitro*. Combining the functional and structural data enables a rational design of mutants with desired characteristics.

5 CONCLUSIONS AND FURTHER PERSPECTIVES

This study discovered important determinants with regard to the function of the Mu transpososome. Structure-function relationships in MuA action that were analyzed by different approaches describe a comprehensive global map of functionally important regions in MuA tetramer. The MuA protein engineering studies that were conducted provided a guide for developing successful directed evolution strategies. Initially, operational ranges of different commonly used random DNA mutagenesis methods and their corresponding mutation spectra were defined. The evaluations generated useful data for selecting a method(s) for the generation of diverse mutant libraries with controlled efficiency of mutagenesis and adjustment of mutation pattern(s). As subsequently established *in vivo* papillation assay provided the means to screen and scrutinize MuA transposase mutants sensitively and quantitatively. As a universal platform, *in vivo* papillation could be used to screen transposase mutants and genes that encode host factors for various mobile elements. It should also be applicable for mechanistic studies to dissect the transposition of novel element families yet to be discovered.

A protein engineering approach that utilized different defined mutagenesis parameters for the generation of high quality MuA mutant libraries following effective *in vivo* papillation screening generated significantly advanced MuA variants for various sets of Mu *in vitro* transposition based applications. The results suggested that MuA transposase did not evolve for maximal activity, even though it is the catalytic component of a virus. The highly active protein variants can be produced by mutating several residues and these substitutions generate excessive hyperactivity due to synergism. Therefore, specific mutation combinations should be found to create the most active protein variant for different applications.

Functional mapping by scanning mutagenesis combined with the sequence alignments of protein homologues yielded comprehensive data on insertion-tolerant versus insertion-intolerant regions in MuA protein family members. In addition, combining the hyperactivity-inducing substitution with the structural data revealed hyperactivity characteristics in Mu transpososome structure. Complementing the analysis with insertion-tolerant versus insertion-intolerant regions allowed further validation of hyperactivity-inducing properties. The results suggest the activity enhancements are associated with DNA binding interactions, active site packing and/or conformational changes, in addition to subunit-subunit interface associations. Several mutations in different protein-protein and protein-DNA interfaces include the catalytic core induced hyperactivity. Thus, these findings expand common knowledge with regard to the overall transpososome function and the architecture of the active site centre of MuA when assembled as a transpososome. All of the DDE-transposases share a common fold of the catalytic core and the active site. Consequently these results should be applicable to other members of this protein family. Attributes of the transpososome revealed in this study may be considered in future studies to improve the transpositional efficiency of a number of different transposable elements.

MuA was found to be a structurally flexible protein and modifiable in its functional features. This potentiates the use of a similar protein engineering strategy for altering target site selection of Mu transposition. In addition, other features revealed in this study are informative for future studies with regard to the function of the Mu transpososome. Combining the functional single substitution and insertion data with those of the structural data sheds light on the rational design of mutants with desired characteristics. Those sites that tolerate insertions could be utilized for different domain insertions including the DNA-binding domain for targeted chromosomal integration, which is a desired feature in gene therapy applications. In general, produced hyperactive MuA transposase variants will improve MuA-based transposon tools and the knowledge gained from this study will enable novel tool design. In addition, an interesting question raised for future research is which of the protomers are critical with regard to particular insertion or amino acid substitution.

ACKNOWLEDGEMENTS

This study has been carried out at the DNA recombination laboratory of the Institute of Biotechnology and Department of Biosciences at the University of Helsinki. I thank my supervisor Professor Harri Savilahti for the interesting research subject and for the excellent facilities in the lab. I am grateful for your guidance and for sharing your thoughts about the transposon world and science in general. I wish to thank for your commitment to this project and the countless hours spent with the preparation of the manuscripts. I highly appreciate the knowledge I have gained during these years in your group.

I would like to thank Professors Mart Saarma and Tomi Mäkelä, the former and current Directors of the Institute of Biotechnology, for providing excellent research facilities. I wish to express my gratitude to Administration Director of the Institute of Biotechnology Arto Halinen for the possibility to complete this work at the Institute of Biotechnology although my supervisor and his research group was transferred to the University of Turku in the middle of my study and also for the financial support on behalf of the Institute of Biotechnology. This work was financially supported by the Finnish Cultural Foundation, the Emil Aaltonen Foundation, the Oskar Öflund Foundation and the University of Helsinki Research Funds.

I thank Professor Jari Valkonen for accepting the invitation to be my opponent, and Professor Jukka Finne for accepting to act as custos at my defence. I thank Docents Suvi Taira and Tuomo Glumoff for their careful and quick review of this thesis and for their insightful comments and suggestions for improvement.

I wish to thank the collaborators who have been involved in the publications of my thesis: Maria Pajunen, Lotta Happonen, Arja Lamberg, Saija Haapa-Paananen, Saija Kiljunen, Mauno Vihinen, Lars Paulin, and Phoebe Rice. Professor Phoebe Rice is acknowledged also for the help and advices in the preparation of PyMol figures. I also want to thank Nisse Kalkkinen for his help with protein purification. Tinde Päivärinta is thanked for the layout of this thesis.

I warmly thank all the past members of the Savilahti group at the Institute of Biotechnology: Maria Pajunen, Pirjo Rahkola, Hilkka Turakainen, Anja Paatero, Sari Tynkkynen, Eini Poussu, Lotta Happonen, Heikki Vilen and Anna-Helena Saariaho. It was nice when there was still a group and not just a lonely graduate student. Thank you for your help, advices, conversations and the fun times. Maria is especially acknowledged for all the scientific and non-scientific help and friendship during these years.

I thank Docent Mikko Frilander for his support and letting to use his premises at the Institute of Biotechnology after Harri's affiliation to the University of Turku. I also want to thank Mikko and his group members for nice conversations at the morning coffee, the fun Christmas parties, and pleasant working environment during the years of being lab-neighbours.

I want to thank all my friends. Special thanks go to Johanna Mantela for sharing the same feelings and frustrations of being graduate student as well as for the fun times. A warm hug goes to Anne Tyrväinen for her valuable friendship since the times at the University of

Jyväskylä. Also, special thanks go to all “Seinis-girls” for keeping the Ostrobothnia spirit alive.

Suuret kiitokset appivanhemmilleni Ilona ja Hannu Rasilalle kaikesta avusta, tuesta, ystävyydestä ja huolenpidosta vuosien varrella. Tämä väitöskirja ei olisi koskaan valmistunut ilman Hemmon hoitolan loistavaa lastenhoitoapua ja kalakeittoa!

Olen syvästi kiitollinen äidilleni elämän mittaisesta kannustuksesta, avusta ja huolenpidosta. Kiitos kaikesta sekä henkisestä että materiaalisesta tuesta. I give warm thanks to my sister Sanna and my brother Sami for all the family events and for your interest in my thesis.

All my love goes to my own family. Thank you my precious little girls Aino and Armi for showing me every day wonders and meaning of life. You are my everything. My dearest thanks go to my best friend and husband Samuli for your endless love and support in different peaks and dips of our lives. Thank you for putting my objectives into the right perspective as well as your patience during these years. Finally, it is time for the post doc era!

Tiina

Espoo, April 2013

References

Abdelhakim, A.H., Oakes, E.C., Sauer, R.T., Baker, T.A., 2008. Unique contacts direct high-priority recognition of the tetrameric Mu transposase-DNA complex by the AAA+ unfoldase ClpX. Mol. Cell 30, 39-50.

Abdelhakim, A.H., Sauer, R.T., Baker, T.A., 2010. The AAA+ ClpX machine unfolds a keystone subunit to remodel the Mu transpososome. Proc. Natl. Acad. Sci. U. S. A. 107, 2437-2442.

Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., Shendure, J., 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. Genome Biol. 11, R119-2010-11-12-r119. Epub 2010 Dec 8.

Adzuma, K., Mizuuchi, K., 1988. Target immunity of Mu transposition reflects a differential distribution of Mu B protein. Cell 53, 257-266.

Adzuma, K., Mizuuchi, K., 1989. Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction. Cell 57, 41-47.

Ahmed, A., 1991. A comparison of intramolecular rearrangements promoted by transposons Tn*5* and Tn*10*. Proc. Biol. Sci. 244, 1-9.

Ahmed, A., 2009. Alternative mechanisms for Tn*5* transposition. PLoS Genet. 5, e1000619.

Akhverdyan, V.Z., Gak, E.R., Tokmakova, I.L., Stoynova, N.V., Yomantas, Y.A., Mashko, S.V., 2011. Application of the bacteriophage Mu-driven system for the integration/amplification of target genes in the chromosomes of engineered Gram-negative bacteria--mini review. Appl. Microbiol. Biotechnol. 91, 857-871.

Akroyd, J.E., Symonds, N., 1983. Evidence for a conservative pathway of transposition of bacteriophage Mu. Nature 303, 84-86.

Aldaz, H., Schuster, E., Baker, T.A., 1996. The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. Cell 85, 257-269.

Allet, B., 1979. Mu insertion duplicates a 5 base pair sequence at the host inserted site. Cell 16, 123-129.

Allison, R.G., Chaconas, G., 1992. Role of the A protein-binding sites in the *in vitro* transposition of Mu DNA. A complex circuit of interactions involving the Mu ends and the transpositional enhancer. J. Biol. Chem. 267, 19963-19970.

Au, T.K., Agrawal, P., Harshey, R.M., 2006. Chromosomal integration mechanism of infecting Mu virion DNA. J. Bacteriol. 188, 1829-1834.

Baker, T.A., Luo, L., 1994. Identification of residues in the Mu transposase essential for catalysis. Proc. Natl. Acad. Sci. U. S. A. 91, 6654-6658.

Baker, T.A., Mizuuchi, K., 1992. DNA-promoted assembly of the active tetramer of the Mu transposase. Genes Dev. 6, 2221-2232.

Baker, T.A., Mizuuchi, M., Mizuuchi, K., 1991. MuB protein allosterically activates strand transfer by the transposase of phage Mu. Cell 65, 1003-1013.

Baker, T.A., Mizuuchi, M., Savilahti, H., Mizuuchi, K., 1993. Division of labor among monomers within the Mu transposase tetramer. Cell 74, 723-733.

Baldwin, A.J., Busse, K., Simm, A.M., Jones, D.D., 2008. Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). Nucleic Acids Res. 36, e77.

Beauregard, A., Curcio, M.J., Belfort, M., 2008. The take and give between retrotransposable elements and their hosts. Annu. Rev. Genet. 42, 587-617.

Bender, J., Kleckner, N., 1992. IS*10* transposase mutations that specifically alter target site recognition. EMBO J. 11, 741-750.

Berg, C.M., Berg, D.E., Groisman, E.A., 1989. Transposable elements and the genetic engineering of bacteria. In: D.E. Berg, M.M. Howe (Eds.), Mobile DNA. American Society for Microbiology, Washington D.C., USA, pp. 879-925.

Berg, C.M., Berg, D.E., 1995. Transposable elements as tool for molecular analyses in bacteria. In: D.J. Sherratt (Ed.), . Oxford University Press Inc., pp. 38-68.

Biemont, C., Vieira, C., 2005. What transposable elements tell us about genome organization and evolution: the case of *Drosophila*. Cytogenet. Genome Res. 110, 25-34.

Biles, B.D., Connolly, B.A., 2004. Low-fidelity *Pyrococcus furiosus* DNA polymerase mutants useful in error-prone PCR. Nucleic Acids Res. 32, e176.

Boeke, J.D., Stoye, J.P., 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: J.M. Coffin, S.H. Hughes, E.H. Varmus (Eds.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, pp. 343-435.

Bolland, S., Kleckner, N., 1996. The three chemical steps of Tn*10*/IS*10* transposition involve repeated utilization of a single active site. Cell 84, 223-233.

Brady, T., Roth, S.L., Malani, N., Wang, G.P., Berry, C.C., Leboulch, P., Hacein-Bey-Abina, S., Cavazzana-Calvo, M., Papapetrou, E.P., Sadelain, M., Savilahti, H., Bushman, F.D., 2011. A method to sequence and quantify DNA integration for monitoring outcome in gene therapy. Nucleic Acids Res. 39, e72.

Brookfield, J.F.Y., 1995. Transposable elements as selfish DNA. In: D.J. Sherratt (Ed.), Oxford University Press Inc., pp. 130-153.

Brookfield, J.F., 2005. The ecology of the genome - mobile DNA elements and their hosts. Nat. Rev. Genet. 6, 128-136.

Bukhari, A.I., 1975. Reversal of mutator phage Mu integration. J. Mol. Biol. 96, 87-99.

Burlingame, R.P., Obukowicz, M.G., Lynn, D.L., Howe, M.M., 1986. Isolation of point mutations in bacteriophage Mu attachment regions cloned in a lambda::mini-Mu phage. Proc. Natl. Acad. Sci. U. S. A. 83, 6012-6016.

Bushman, F., 2002. Lateral DNA Transfer: Mechanisms and Consequences. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Butterfield, Y.S., Marra, M.A., Asano, J.K., Chan, S.Y., Guin, R., Krzywinski, M.I., Lee, S.S., MacDonald, K.W., Mathewson, C.A., Olson, T.E., Pandoh, P.K., Prabhu, A.L., Schnerch, A., Skalska, U., Smailus, D.E., Stott, J.M., Tsai, M.I., Yang, G.S., Zuyderduyn, S.D., Schein, J.E., Jones, S.J., 2002. An efficient strategy for large-scale high-throughput transposon-mediated sequencing of cDNA clones. Nucleic Acids Res. 30, 2460-2468.

Cadwell, R.C., Joyce, G.F., 1992. Randomization of genes by PCR mutagenesis. PCR Methods Appl. 2, 28-33.

Campbell, A., 2002. Eubacterial genomes. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambowitz (Eds.), Mobile DNA II. ASM Press, Washington, D.C., pp. 1024-1039.

Camps, M., Naukkarinen, J., Johnson, B.P., Loeb, L.A., 2003. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. Proc. Natl. Acad. Sci. U. S. A. 100, 9727-9732.

Chaconas, G., Harshey, R.M., 2002. Transposition of Phage Mu DNA. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambowitz (Eds.), Mobile DNA II. ASM Press, Washington, DC, pp. 384-402.

Chaconas, G., Harshey, R.M., Sarvetnick, N., Bukhari, A.I., 1981. Predominant end-products of prophage Mu DNA transposition during the lytic cycle are replicon fusions. J. Mol. Biol. 150, 341-359.

Chaconas, G., Kennedy, D.L., Evans, D., 1983. Predominant integration end products of infecting bacteriophage Mu DNA are simple insertions with no preference for integration of either Mu DNA strand. Virology 128, 48-59.
Chandler, M., Mahillon, J., 2002. Insertion sequences revisited. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambowitz (Eds.), . ASM Press, Washington, DC, pp. 305-366.
Cherepanov, P., Maertens, G.N., Hare, S., 2011. Structural insights into the retroviral DNA integration apparatus. Curr. Opin. Struct. Biol. 21, 249-256.
Choi, W., Harshey, R.M., 2010. DNA repair by the cryptic endonuclease activity of Mu transposase. Proc. Natl. Acad. Sci. U. S. A. 107, 10014-10019.
Claeys Bouuaert, C., Liu, D., Chalmers, R., 2011. A simple topological filter in a eukaryotic transposon as a mechanism to suppress genome instability. Mol. Cell. Biol. 31, 317-327.
Cline, J., Braman, J.C., Hogrefe, H.H., 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. 24, 3546-3551.
Clubb, R.T., Mizuuchi, M., Huth, J.R., Omichinski, J.G., Savilahti, H., Mizuuchi, K., Clore, G.M., Gronenborn, A.M., 1996. The wing of the enhancer-binding domain of Mu phage transposase is flexible and is essential for efficient transposition. Proc. Natl. Acad. Sci. U. S. A. 93, 1146-1150.
Clubb, R.T., Schumacher, S., Mizuuchi, K., Gronenborn, A.M., Clore, G.M., 1997. Solution structure of the Iγ subdomain of the Mu end DNA-binding domain of phage Mu transposase. J. Mol. Biol. 273, 19-25.
Copeland, N.G., Jenkins, N.A., 2010. Harnessing transposons for cancer gene discovery. Nat. Rev. Cancer. 10, 696-706.
Craig, N., 2002. Tn*7*. In: L. Craig Nancy, R. Craigie, M. Gellert, M. Lambowitz-Alan (Eds.), Mobile DNA II. ASM Press, Washington, DC, pp. 423-456.
Craig, N.L., 1995. Unity in transposition reactions. Science 270, 253-254.
Craig, N.L., 1997. Target site selection in transposition. Annu. Rev. Biochem. 66, 437-474.
Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A.M. (Eds.), 2002. Mobile DNA II. ASM Press, Washington, DC.
Craigie, R., 1996. Quality control in Mu DNA transposition. Cell 85, 137-140.
Craigie, R., Arndt-Jovin, D.J., Mizuuchi, K., 1985. A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. Proc. Natl. Acad. Sci. U. S. A. 82, 7570-7574.
Craigie, R., Mizuuchi, K., 1985. Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. Cell 41, 867-876.
Craigie, R., Mizuuchi, K., 1986. Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. Cell 45, 793-800.
Craigie, R., Mizuuchi, K., 1987. Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. Cell 51, 493-501.
Craigie, R., Mizuuchi, M., Mizuuchi, K., 1984. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. Cell 39, 387-394.
Curcio, M.J., Derbyshire, K.M., 2003. The outs and ins of transposition: from Mu to kangaroo. Nat. Rev. Mol. Cell Biol. 4, 865-877.
Dalby, P.A., 2011. Strategy and success for the directed evolution of enzymes. Curr. Opin. Struct. Biol. 21, 473-480.
Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., Rayment, I., 2000. Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate. Science 289, 77-85.
Derbyshire, K.M., Grindley, N.D., 1996. *Cis* preference of the IS*903* transposase is mediated by a combination of transposase instability and inefficient translation. Mol. Microbiol. 21, 1261-1272.

Doherty, J.E., Huye, L.E., Yusa, K., Zhou, L., Craig, N.L., Wilson, M.H., 2012. Hyperactive *piggyBac* gene transfer in human cells and *in vivo*. Hum. Gene Ther. 23, 311-320.

Dyda, F., Hickman, A., Burgess, 2008. DNA transposases. In: P.A. Rice, C.C. Correll (Eds.), Protein_Nucleic Acid Interactions:Structural Biology. Royal Society of Chemistry, Cambridge, UK, pp. 270-302.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R., Davies, D.R., 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science 266, 1981-1986.

Echols, H., 1986. Multiple DNA-protein interactions governing high-precision DNA transactions. Science 233, 1050-1056.

Eckert, K.A., Kunkel, T.A., 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Res. 18, 3739-3744.

Edwards, W.R., Busse, K., Allemann, R.K., Jones, D.D., 2008. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. Nucleic Acids Res. 36, e78.

Emond, S., Mondon, P., Pizzut-Serin, S., Douchy, L., Crozet, F., Bouayadi, K., Kharrat, H., Potocki-Veronese, G., Monsan, P., Remaud-Simeon, M., 2008. A novel random mutagenesis approach using human mutagenic DNA polymerases to generate enzyme variant libraries. Protein Eng. Des. Sel. 21, 267-274.

Evgen'ev, M.B., Arkhipova, I.R., 2005. Penelope-like elements--a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenet. Genome Res. 110, 510-521.

Evgen'ev, M.B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D.H., Corces, V.G., 1997. *Penelope*, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. Proc. Natl. Acad. Sci. U. S. A. 94, 196-201.

Fedoroff, N., 2002. Contol of Mobile DNA. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambowitz (Eds.), Mobile DNA II. ASM Press, Washington, DC, pp. 997-1007.

Feng, X., Colloms, S.D., 2007. *In vitro* transposition of ISY*100*, a bacterial insertion sequence belonging to the Tc1/*mariner* family. Mol. Microbiol. 65, 1432-1443.

Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9, 397-405.

Feschotte, C., Jiang, N., Wessler, S.R., 2002. Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. 3, 329-341.

Feschotte, C., Pritham, E.J., 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41, 331-368.

Filee, J., Siguier, P., Chandler, M., 2007. Insertion sequence diversity in archaea. Microbiol. Mol. Biol. Rev. 71, 121-157.

Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. Trends Genet. 5, 103-107.

Finnegan, D.J., 1990. Transposable elements and DNA transposition in eukaryotes. Curr. Opin. Cell Biol. 2, 471-477.

Fujii, R., Kitaoka, M., Hayashi, K., 2004. One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Res. 32, e145.

Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S.J., Craig, N.L., 2010. DNA transposon *Hermes* inserts into DNA in nucleosome-free regions in vivo. Proc. Natl. Acad. Sci. U. S. A. 107, 21966-21972.

Ge, J., Lou, Z., Harshey, R.M., 2010. Immunity of replicating Mu to self-integration: a novel mechanism employing MuB protein. Mob DNA 1, 8-8753-1-8.

Gellert, M., 2002. V(D)J recombination: RAG proteins, repair factors, and regulation. Annu. Rev. Biochem. 71, 101-132.

Gertz, J., Varley, K.E., Davis, N.S., Baas, B.J., Goryshin, I.Y., Vaidyanathan, R., Kuersten, S., Myers, R.M., 2012. Transposase mediated construction of RNA-seq libraries. Genome Res. 22, 134-141.

Goldhaber-Gordon, I., Early, M.H., Gray, M.K., Baker, T.A., 2002a. Sequence and positional requirements for DNA sites in a Mu transpososome. J. Biol. Chem. 277, 7703-7712.

Goldhaber-Gordon, I., Williams, T.L., Baker, T.A., 2002b. DNA recognition sites activate MuA transposase to perform transposition of non-Mu DNA. J. Biol. Chem. 277, 7694-7702.

Goodier, J.L., Kazazian, H.H.,Jr, 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 135, 23-35.

Goodwin, T.J., Poulter, R.T., 2001. The DIRS1 group of retrotransposons. Mol. Biol. Evol. 18, 2067-2082.

Goodwin, T.J., Poulter, R.T., 2004. A new group of tyrosine recombinase-encoding retrotransposons. Mol. Biol. Evol. 21, 746-759.

Goryshin, I.Y., Reznikoff, W.S., 1998. Tn*5 in vitro* transposition. J. Biol. Chem. 273, 7367-7374.

Grabundzija, I., Irgang, M., Mates, L., Belay, E., Matrai, J., Gogol-Doring, A., Kawakami, K., Chen, W., Ruiz, P., Chuah, M.K., VandenDriessche, T., Izsvák, Z., Ivics, Z., 2010. Comparative analysis of transposable element vector systems in human cells. Mol. Ther. 18, 1200-1209.

Gray, Y.H., 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet. 16, 461-468.

Greene, E.C., Mizuuchi, K., 2002a. Direct observation of single MuB polymers: evidence for a DNA-dependent conformational change for generating an active target complex. Mol. Cell 9, 1079-1089.

Greene, E.C., Mizuuchi, K., 2002b. Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. EMBO J. 21, 1477-1486.

Greene, E.C., Mizuuchi, K., 2002c. Target immunity during Mu DNA transposition. Transpososome assembly and DNA looping enhance MuA-mediated disassembly of the MuB target complex. Mol. Cell 10, 1367-1378.

Greener, A., Callahan, M., Jerpseth, B., 1996. An efficient random mutagenesis technique using an *E. coli* mutator strain. Methods Mol. Biol. 57, 375-385.

Gregory, T.R., 2005. Synergy between sequence and size in large-scale genomics. Nat. Rev. Genet. 6, 699-708.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., Gelbart, W.M., 2000. An Introduction to Genetic Analysis. W. H. Freeman And Company, New York.

Grindley, N.D., Whiteson, K.L., Rice, P.A., 2006. Mechanisms of site-specific recombination. Annu. Rev. Biochem. 75, 567-605.

Grindley, N.D.F., 2002. The movement of Tn*3*-like elements: transposition and cointegrate resolution. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambowitz (Eds.), . ASM Press, Washington, DC, pp. 272-302.

Groenen, M.A., van de Putte, P., 1986. Analysis of the ends of bacteriophage Mu using site-directed mutagenesis. J. Mol. Biol. 189, 597-602.

Gueguen, E., Rousseau, P., Duval-Valentin, G., Chandler, M., 2005. The transpososome: control of transposition at the level of catalysis. Trends Microbiol. 13, 543-549.

Haapa, S., Suomalainen, S., Eerikainen, S., Airaksinen, M., Paulin, L., Savilahti, H., 1999a. An efficient DNA sequencing strategy based on the bacteriophage Mu *in vitro* DNA transposition reaction. Genome Res. 9, 308-315.

Haapa, S., Taira, S., Heikkinen, E., Savilahti, H., 1999b. An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications. Nucleic Acids Res. 27, 2777-2784.

Haapa-Paananen, S., Rita, H., Savilahti, H., 2002. DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection *in vitro* . J. Biol. Chem. 277, 2843-2851.

Hackett, P.B., Largaespada, D.A., Cooper, L.J., 2010. A transposon and transposase system for human application. Mol. Ther. 18, 674-683.

Halling, S.M., Kleckner, N., 1982. A symmetrical six-base-pair target site sequence determines Tn*10* insertion specificity. Cell 28, 155-163.

Hamer, L., DeZwaan, T.M., Montenegro-Chamorro, M.V., Frank, S.A., Hamer, J.E., 2001. Recent advances in large-scale transposon mutagenesis. Curr. Opin. Chem. Biol. 5, 67-73.

Han, Y.W., Mizuuchi, K., 2010. Phage Mu transposition immunity: protein pattern formation along DNA by a diffusion-ratchet mechanism. Mol. Cell 39, 48-58.

Haniford, D.B., 2006. Transpososome dynamics and regulation in Tn*10* transposition. Crit. Rev. Biochem. Mol. Biol. 41, 407-424.

Hare, S., Gupta, S.S., Valkov, E., Engelman, A., Cherepanov, P., 2010a. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature 464, 232-236.

Hare, S., Vos, A.M., Clayton, R.F., Thuring, J.W., Cummings, M.D., Cherepanov, P., 2010b. Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. Proc. Natl. Acad. Sci. U. S. A. 107, 20057-20062.

Haren, L., Polard, P., Ton-Hoang, B., Chandler, M., 1998. Multiple oligomerisation domains in the IS*911* transposase: a leucine zipper motif is essential for activity. J. Mol. Biol. 283, 29-41.

Haren, L., Ton-Hoang, B., Chandler, M., 1999. Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. 53, 245-281.

Harshey, R.M., 1984. Transposition without duplication of infecting bacteriophage Mu DNA. Nature 311, 580-581.

Harshey, R.M., 2012. The Mu story: how a maverick phage moved the field forward. Mob DNA 3, 21-8753-3-21.

Harshey, R.M., Getzoff, E.D., Baldwin, D.L., Miller, J.L., Chaconas, G., 1985. Primary structure of phage Mu transposase: homology to Mu repressor. Proc. Natl. Acad. Sci. U. S. A. 82, 7676-7680.

Harshey, R.M., Jayaram, M., 2006. The Mu transpososome through a topological lens. Crit. Rev. Biochem. Mol. Biol. 41, 387-405.

Hashey, R.M., Cuneo, S.D., 1986. Carboxyl-terminal mutants of phage Mu transposase. J. Genet 65, 159-174.

Hayes, F., 2003. Transposon-based strategies for microbial functional genomics and proteomics. Annu. Rev. Genet. 37, 3-29.

Hayes, F., Hallet, B., 2000. Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks. Trends Microbiol. 8, 571-577.

Heinz, D.W., Baase, W.A., Zhang, X.J., Blaber, M., Dahlquist, F.W., Matthews, B.W., 1994. Accommodation of amino acid insertions in an alpha-helix of T4 lysozyme. Structural and thermodynamic analysis. J. Mol. Biol. 236, 869-886.

Hickman, A.B., Chandler, M., Dyda, F., 2010. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. Crit. Rev. Biochem. Mol. Biol. 45, 50-69.

Hickman, A.B., Perez, Z.N., Zhou, L., Musingarimi, P., Ghirlando, R., Hinshaw, J.E., Craig, N.L., Dyda, F., 2005. Molecular architecture of a eukaryotic DNA transposase. Nat. Struct. Mol. Biol. 12, 715-721.

Hoeller, B.M., Reiter, B., Abad, S., Graze, I., Glieder, A., 2008. Random tag insertions by Transposon Integration mediated Mutagenesis (TIM). J. Microbiol. Methods 75, 251-257.

Hu, W.Y., Derbyshire, K.M., 1998. Target choice and orientation preference of the insertion sequence IS*903*. J. Bacteriol. 180, 3039-3048.

Huisman, O., Kleckner, N., 1987. A new generalizable test for detection of mutations affecting Tn*10* transposition. Genetics 116, 185-189.

Ivics, Z., Izsvák, Z., 2010. The expanding universe of transposon technologies for gene and cell engineering. Mob DNA 1, 25.

Ivics, Z., Li, M.A., Mates, L., Boeke, J.D., Nagy, A., Bradley, A., Izsvák, Z., 2009. Transposon-mediated genome manipulation in vertebrates. Nat. Methods 6, 415-422.

Izsvák, Z., Khare, D., Behlke, J., Heinemann, U., Plasterk, R.H., Ivics, Z., 2002. Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. J. Biol. Chem. 277, 34581-34588.

Jäckel, C., Kast, P., Hilvert, D., 2008. Protein Design by Directed Evolution. The Annual Reviw of Biophysics 37, 153-173.

Jang, S., Sandler, S.J., Harshey, R.M., 2012. Mu insertions are repaired by the double-strand break repair pathway of *Escherichia coli*. PLoS Genet. 8, e1002642.

Jiang, H., Harshey, R.M., 2001. The Mu enhancer is functionally asymmetric both *in cis* and *in trans*. Topological selectivity of Mu transposition is enhancer-independent. J. Biol. Chem. 276, 4373-4381.

Jones, D.D., 2005. Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 beta-lactamase to an amino acid deletion. Nucleic Acids Res. 33, e80.

Jones, J.M., Welty, D.J., Nakai, H., 1998. Versatile action of *Escherichia coli* ClpXP as protease or molecular chaperone for bacteriophage Mu transposition. J. Biol. Chem. 273, 459-465.

Judson, N., Mekalanos, J.J., 2000. Transposon-based approaches to identify essential bacterial genes. Trends Microbiol. 8, 521-526.

Jukkola, T., Trokovic, R., Maj, P., Lamberg, A., Mankoo, B., Pachnis, V., Savilahti, H., Partanen, J., 2005. Meox1Cre: a mouse line expressing Cre recombinase in somitic mesoderm. Genesis 43, 148-153.

Junop, M.S., Haniford, D.B., 1997. Factors responsible for target site selection in Tn*10* transposition: a role for the DDE motif in target DNA capture. EMBO J. 16, 2646-2655.

Jurka, J., Kapitonov, V.V., Kohany, O., Jurka, M.V., 2007. Repetitive sequences in complex genomes: structure and evolution. Annu. Rev. Genomics Hum. Genet. 8, 241-259.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462-467.

Kahlig, K.M., Saridey, S.K., Kaja, A., Daniels, M.A., George, A.L.,Jr, Wilson, M.H., 2010. Multiplexed transposon-mediated stable gene transfer in human cells. Proc. Natl. Acad. Sci. U. S. A. 107, 1343-1348.

Kahmann, R., Kamp, D., 1979. Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. Nature 280, 247-250.

Kaiser, K., Sentry, J.W., Finnegan, D.J., 1995. Eukaryotic transposable elements as tools to study gene structure and function. In: D.J. Sherratt (Ed.), . Oxford University Press Inc., pp. 69-100.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., Morikawa, K., 1990. Three-dimensional structure of ribonuclease H from *E. coli*. Nature 347, 306-309.

Kaur, J., Sharma, R., 2006. Directed evolution: an approach to engineer enzymes. Crit. Rev. Biotechnol. 26, 165-199.

Kekarainen, T., Savilahti, H., Valkonen, J.P., 2002. Functional genomics on potato virus A: virus genome-wide map of sites essential for virus propagation. Genome Res. 12, 584-594.

Kennedy, A.K., Haniford, D.B., Mizuuchi, K., 2000. Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. Cell 101, 295-305.

Kidwell, M.G., Lisch, D., 1997. Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. U. S. A. 94, 7704-7711.

Kidwell, M.G., Lisch, D.R., 2001. Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution 55, 1-24.

Kiljunen, S., Vilen, H., Pajunen, M., Savilahti, H., Skurnik, M., 2005. Nonessential genes of phage phiYeO3-12 include genes involved in adaptation to growth on Yersinia enterocolitica serotype O:3. J. Bacteriol. 187, 1405-1414.

Kim, K., Harshey, R.M., 1995. Mutational analysis of the att DNA-binding domain of phage Mu transposase. Nucleic Acids Res. 23, 3937-3943.

Kim, K., Namgoong, S.Y., Jayaram, M., Harshey, R.M., 1995. Step-arrest mutants of phage Mu transposase. Implications in DNA-protein assembly, Mu end cleavage, and strand transfer. J. Biol. Chem. 270, 1472-1479.

Kim, Y.C., Morrison, S.L., 2009. N-terminal domain-deleted Mu transposase exhibits increased transposition activity with low target site preference in modified buffers. J. Mol. Microbiol. Biotechnol. 17, 30-40.

Kobryn, K., Lavoie, B.D., Chaconas, G., 1999. Supercoiling-dependent site-specific binding of HU to naked Mu DNA. J. Mol. Biol. 289, 777-784.

Kopsidas, G., Carman, R.K., Stutt, E.L., Raicevic, A., Roberts, A.S., Siomos, M.A., Dobric, N., Pontes-Braz, L., Coia, G., 2007. RNA mutagenesis yields highly diverse mRNA libraries for *in vitro* protein evolution. BMC Biotechnol. 7, 18.

Krause, H.M., Higgins, N.P., 1986. Positive and negative regulation of the Mu operator by Mu repressor and *Escherichia coli* integration host factor. J. Biol. Chem. 261, 3744-3752.

Krebs, M.P., Reznikoff, W.S., 1988. Use of a Tn*5* derivative that creates *lacZ* translational fusions to obtain a transposition mutant. Gene 63, 277-285.

Krementsova, E., Giffin, M.J., Pincus, D., Baker, T.A., 1998. Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. J. Biol. Chem. 273, 31358-31365.

Kruklitis, R., Nakai, H., 1994. Participation of the bacteriophage Mu A protein and host factors in the initiation of Mu DNA synthesis *in vitro*. J. Biol. Chem. 269, 16469-16477.

Kruklitis, R., Welty, D.J., Nakai, H., 1996. ClpX protein of *Escherichia coli* activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. EMBO J. 15, 935-944.

Krupovic, M., Vilen, H., Bamford, J.K., Kivela, H.M., Aalto, J.M., Savilahti, H., Bamford, D.H., 2006. Genome characterization of lipid-containing marine bacteriophage PM2 by transposon insertion mutagenesis. J. Virol. 80, 9270-9278.

Kuduvalli, P.N., Rao, J.E., Craig, N.L., 2001. Target DNA structure plays a critical role in Tn*7* transposition. EMBO J. 20, 924-932.

Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P., Skalka, A.M., 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. 12, 2331-2338.

Kumar, A., Singh, S., 2012. Directed evolution: tailoring biocatalysts for industrial applications. Crit. Rev. Biotechnol.

Kuo, C.F., Zou, A.H., Jayaram, M., Getzoff, E., Harshey, R., 1991. DNA-protein complexes during attachment-site synapsis in Mu DNA transposition. EMBO J. 10, 1585-1591.

Lai, Y.P., Huang, J., Wang, L.F., Li, J., Wu, Z.R., 2004. A new approach to random mutagenesis *in vitro*. Biotechnol. Bioeng. 86, 622-627.

Lamberg, A., Nieminen, S., Qiao, M., Savilahti, H., 2002. Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of *in vitro*-assembled DNA transposition complexes of bacteriophage Mu. Appl. Environ. Microbiol. 68, 705-712.

Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J., Robertson, H.M., 1999. Hyperactive transposase mutants of the *Himar1* mariner transposon. Proc. Natl. Acad. Sci. U. S. A. 96, 11428-11433.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J, et al. International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Laurent, L.C., Olsen, M.N., Crowley, R.A., Savilahti, H., Brown, P.O., 2000. Functional characterization of the human immunodeficiency virus type 1 genome by genetic footprinting. J. Virol. 74, 2760-2769.

Lavoie, B.D., Chan, B.S., Allison, R.G., Chaconas, G., 1991. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. EMBO J. 10, 3051-3059.

Lavoie, B.D., Shaw, G.S., Millner, A., Chaconas, G., 1996. Anatomy of a flexer-DNA complex inside a higher-order transposition intermediate. Cell 85, 761-771.

Lee, I., Harshey, R.M., 2001. Importance of the conserved CA dinucleotide at Mu termini. J. Mol. Biol. 314, 433-444.

Lee, I., Harshey, R.M., 2003a. The conserved CA/TG motif at Mu termini: T specifies stable transpososome assembly. J. Mol. Biol. 330, 261-275.

Lee, I., Harshey, R.M., 2003b. Patterns of sequence conservation at termini of long terminal repeat (LTR) retrotransposons and DNA transposons in the human genome: lessons from phage Mu. Nucleic Acids Res. 31, 4531-4540.

Leung, D.W., Chen, E., Goeddel, D.V., 1989. A Method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Techniques 1, 11-15.

Leung, P.C., Harshey, R.M., 1991. Two mutations of phage Mu transposase that affect strand transfer or interactions with B protein lie in distinct polypeptide domains. J. Mol. Biol. 219, 189-199.

Leung, P.C., Teplow, D.B., Harshey, R.M., 1989. Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. Nature 338, 656-658.

Levchenko, I., Luo, L., Baker, T.A., 1995. Disassembly of the Mu transposase tetramer by the ClpX chaperone. Genes Dev. 9, 2399-2408.

Levchenko, I., Yamauchi, M., Baker, T.A., 1997. ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. Genes Dev. 11, 1561-1572.

Liebart, J.C., Ghelardini, P., Paolozzi, L., 1982. Conservative integration of bacteriophage Mu DNA into pBR322 plasmid. Proc. Natl. Acad. Sci. U. S. A. 79, 4362-4366.

Lipkow, K., Buisine, N., Lampe, D.J., Chalmers, R., 2004. Early intermediates of *Mariner* transposition: catalysis without synapsis of the transposon ends suggests a novel architecture of the synaptic complex. Mol. Cell. Biol. 24, 8301-8311.

Lisch, D., 2012. How important are transposons for plant evolution?. Nat. Rev. Genet. 14, 49-61.

Liu, S., Yeh, C.T., Ji, T., Ying, K., Wu, H., Tang, H.M., Fu, Y., Nettleton, D., Schnable, P.S., 2009. Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genet. 5, e1000733.

Loot, C., Turlan, C., Rousseau, P., Ton-Hoang, B., Chandler, M., 2002. A target specificity switch in IS*911* transposition: the role of the OrfA protein. EMBO J. 21, 4172-4182.

Lu, F., Craig, N.L., 2000. Isolation and characterization of Tn*7* transposase gain-of-function mutants: a model for transposase activation. EMBO J. 19, 3446-3457.

Maertens, G.N., Hare, S., Cherepanov, P., 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature 468, 326-329.

Mahillon, J., Chandler, M., 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. 62, 725-774.

Makris, J.C., Nordmann, P.L., Reznikoff, W.S., 1988. Mutational analysis of insertion sequence *50* (IS*50*) and transposon *5* (Tn*5*) ends. Proc. Natl. Acad. Sci. U. S. A. 85, 2224-2228.

Manna, D., Wang, X., Higgins, N.P., 2001. Mu and IS*1* transpositions exhibit strong orientation bias at the *Escherichia coli bgl* locus. J. Bacteriol. 183, 3328-3335.

Mariconda, S., Namgoong, S.Y., Yoon, K.H., Jiang, H., Harshey, R.M., 2000. Domain III function of Mu transposase analysed by directed placement of subunits within the transpososome. J. Biosci. 25, 347-360.

Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., VandenDriessche, T., Ivics, Z., Izsvák, Z., 2009. Molecular evolution of a novel hyperactive *Sleeping Beauty* transposase enables robust stable gene transfer in vertebrates. Nat. Genet. 41, 753-761.

Maxwell, A., Craigie, R., Mizuuchi, K., 1987. B protein of bacteriophage Mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. Proc. Natl. Acad. Sci. U. S. A. 84, 699-703.

May, E.W., Craig, N.L., 1996. Switching from cut-and-paste to replicative Tn*7* transposition. Science 272, 401-404.

McClintock Barbara, 1987. The Discovery and Characterization of Transposable Elements: the Collected Papers of Barbara McClintock. Garland, Mew York, NY.

Miller, J.H., 1992. A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria. Cold Spring Harbor Laboratory Press, New York.

Miller, J.H., 1996. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. Annu. Rev. Microbiol. 50, 625-643.

Miller, J.H., 1998. Mutators in *Escherichia coli* . Mutat. Res. 409, 99-106.

Miskey, C., Izsvák, Z., Kawakami, K., Ivics, Z., 2005. DNA transposons in vertebrate functional genomics. Cell Mol. Life Sci. 62, 629-641.

Miyazaki, K., Arnold, F.H., 1999. Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. J. Mol. Evol. 49, 716-720.

Mizuuchi, K., 1983. *In vitro* transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. Cell 35, 785-794.

Mizuuchi, K., 1984. Mechanism of transposition of bacteriophage Mu: polarity of the strand transfer reaction at the initiation of transposition. Cell 39, 395-404.

Mizuuchi, K., 1992a. Polynucleotidyl transfer reactions in transpositional DNA recombination. J. Biol. Chem. 267, 21273-21276.

Mizuuchi, K., 1992b. Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. 61, 1011-1051.

Mizuuchi, K., Adzuma, K., 1991. Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. Cell 66, 129-140.

Mizuuchi, K., Baker, T.A., 2002. Chemical mechanisms for mobilizing DNA. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambowitz (Eds.), Mobile DNA II. ASM Press, Washington, DC, pp. 12-23.

Mizuuchi, K., Craigie, R., 1986. Mechanism of bacteriophage Mu transposition. Annu. Rev. Genet. 20, 385-429.
Mizuuchi, M., Baker, T.A., Mizuuchi, K., 1991. DNase protection analysis of the stable synaptic complexes involved in Mu transposition. Proc. Natl. Acad. Sci. U. S. A. 88, 9031-9035.
Mizuuchi, M., Baker, T.A., Mizuuchi, K., 1992. Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. Cell 70, 303-311.
Mizuuchi, M., Mizuuchi, K., 1989. Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. Cell 58, 399-408.
Mizuuchi, M., Mizuuchi, K., 1993. Target site selection in transposition of phage Mu. Cold Spring Harb. Symp. Quant. Biol. 58, 515-523.
Montaňo, S.P., Pigli, Y.Z., Rice, P.A., 2012. The Mu transpososome structure sheds light on DDE recombinase evolution. Nature 491, 413-417.
Montaňo, S.P., Rice, P.A., 2011. Moving DNA around: DNA transposition and retroviral integration. Curr. Opin. Struct. Biol. 21, 370-378.
Morgan, G.J., Hatfull, G.F., Casjens, S., Hendrix, R.W., 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. J. Mol. Biol. 317, 337-359.
Muotri, A.R., Marchetto, M.C., Coufal, N.G., Gage, F.H., 2007. The necessary junk: new functions for transposable elements. Hum. Mol. Genet. 16 Spec No. 2, R159-67.
Murakami, H., Hohsaka, T., Sisido, M., 2003. Random insertion and deletion mutagenesis. In: F.A. Arnold, G. Georgiou (Eds.), Methods in Molecular Biology. Humana Press, Totowa, New Jersey, pp. 53-64.
Myers, R.M., Lerman, L.S., Maniatis, T., 1985. A general method for saturation mutagenesis of cloned DNA fragments. Science 229, 242-247.
Nagy, Z., Chandler, M., 2004. Regulation of transposition in bacteria. Res. Microbiol. 155, 387-398.
Naigamwalla, D.Z., Chaconas, G., 1997. A new set of Mu DNA transposition intermediates: alternate pathways of target capture preceding strand transfer. EMBO J. 16, 5227-5234.
Nakai, H., Doseeva, V., Jones, J.M., 2001. Handoff from recombinase to replisome: insights from transposition. Proc. Natl. Acad. Sci. U. S. A. 98, 8247-8254.
Nakayama, C., Teplow, D.B., Harshey, R.M., 1987. Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. Proc. Natl. Acad. Sci. U. S. A. 84, 1809-1813.
Namgoong, S.Y., Harshey, R.M., 1998. The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. EMBO J. 17, 3775-3785.
Namgoong, S.Y., Jayaram, M., Kim, K., Harshey, R.M., 1994. DNA-protein cooperativity in the assembly and stabilization of Mu strand transfer complex. Relevance of DNA phasing and att site cleavage. J. Mol. Biol. 238, 514-527.
Namgoong, S.Y., Sankaralingam, S., Harshey, R.M., 1998. Altering the DNA-binding specificity of Mu transposase *in vitro*. Nucleic Acids Res. 26, 3521-3527.
Naumann, T.A., Reznikoff, W.S., 2002. Tn*5* transposase with an altered specificity for transposon ends. J. Bacteriol. 184, 233-240.
Nesmelova, I.V., Hackett, P.B., 2010. DDE transposases: Structural similarity and diversity. Adv. Drug Deliv. Rev. 62, 1187-1195.
Neylon, C., 2004. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Res. 32, 1448-1459.

Ni, J., Clark, K.J., Fahrenkrug, S.C., Ekker, S.C., 2008. Transposon tools hopping in vertebrates. Brief Funct. Genomic Proteomic 7, 444-453.

Nieminen, M., Tuuri, T., Savilahti, H., 2010. Genetic recombination pathways and their application for genome modification of human embryonic stem cells. Exp. Cell Res. 316, 2578-2586.

Nowotny, M., 2009. Retroviral integrase superfamily: the structural perspective. EMBO Rep. 10, 144-151.

Nowotny, M., Gaidamakov, S.A., Crouch, R.J., Yang, W., 2005. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell 121, 1005-1016.

Nowotny, M., Yang, W., 2006. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. EMBO J. 25, 1924-1933.

Oliver, K.R., Greene, W.K., 2009. Transposable elements: powerful facilitators of evolution. Bioessays 31, 703-714.

O'Malley, J., Woltjen, K., Kaji, K., 2009. New strategies to generate induced pluripotent stem cells. Curr. Opin. Biotechnol. 20, 516-521.

Orsini, L., Pajunen, M., Hanski, I., Savilahti, H., 2007. SNP discovery by mismatch-targeting of Mu transposition. Nucleic Acids Res. 35, e44.

Paatero, A.O., Turakainen, H., Happonen, L.J., Olsson, C., Palomaki, T., Pajunen, M.I., Meng, X., Otonkoski, T., Tuuri, T., Berry, C., Malani, N., Frilander, M.J., Bushman, F.D., Savilahti, H., 2008. Bacteriophage Mu integration in yeast and mammalian genomes. Nucleic Acids Res. 36, e148.

Pajunen, M., Turakainen, H., Poussu, E., Peränen, J., Vihinen, M., Savilahti, H., 2007. High-precision mapping of protein protein interfaces: an integrated genetic strategy combining *en masse* mutagenesis and DNA-level parallel analysis on a yeast two-hybrid platform. Nucleic Acids Res. 35, e103.

Pajunen, M.I., Pulliainen, A.T., Finne, J., Savilahti, H., 2005. Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes. Microbiology 151, 1209-1218.

Parks, A.R., Li, Z., Shi, Q., Owens, R.M., Jin, M.M., Peters, J.E., 2009. Transposition into replicating DNA occurs through interaction with the processivity factor. Cell 138, 685-695.

Patel, P.H., Kawate, H., Adman, E., Ashbach, M., Loeb, L.A., 2001. A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. J. Biol. Chem. 276, 5044-5051.

Pathania, S., Jayaram, M., Harshey, R.M., 2002. Path of DNA within the Mu transpososome. Transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils. Cell 109, 425-436.

Pathania, S., Jayaram, M., Harshey, R.M., 2003. A unique right end-enhancer complex precedes synapsis of Mu ends: the enhancer is sequestered within the transpososome throughout transposition. EMBO J. 22, 3725-3736.

Patrick, W.M., Firth, A.E., Blackburn, J.M., 2003. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein Eng. 16, 451-457.

Peters, J.E., Craig, N.L., 2001. Tn7: smarter than we thought. Nat. Rev. Mol. Cell Biol. 2, 806-814.

Plasterk, R.H.A., 1995. Mechanisms of DNA transposition. In: D.J. Sherratt (Ed.), Mobile Genetic Elements. Oxford University Press Inc., pp. 18-37.

Plasterk, R.H., Izsvák, Z., Ivics, Z., 1999. Resident aliens: the Tc1/*mariner* superfamily of transposable elements. Trends Genet. 15, 326-332.

Polard, P., Chandler, M., 1995. Bacterial transposases and retroviral integrases. Mol. Microbiol. 15, 13-23.

Poussu, E., Jäntti, J., Savilahti, H., 2005. A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: mapping of the Sec1p binding domain in yeast Mso1p by a Mu *in vitro* transposition-based approach. Nucleic Acids Res. 33, e104.

Poussu, E., Vihinen, M., Paulin, L., Savilahti, H., 2004. Probing the α-complementing domain of *E. coli* β-galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu *in vitro* DNA transposition. Proteins 54, 681-692.

Pribil, P.A., Haniford, D.B., 2003. Target DNA bending is an important specificity determinant in target site selection in Tn10 transposition. J. Mol. Biol. 330, 247-259.

Pritham, E.J., 2009. Transposable elements and factors influencing their success in eukaryotes. J. Hered. 100, 648-655.

Rao, J.E., Miller, P.S., Craig, N.L., 2000. Recognition of triple-helical DNA structures by transposon Tn*7*. Proc. Natl. Acad. Sci. U. S. A. 97, 3936-3941.

Rebollo, R., Horard, B., Hubert, B., Vieira, C., 2010. Jumping genes and epigenetics: Towards new species. Gene 454, 1-7.

Reznikoff, W.S., 2002. Tn*5* transposition. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambowitz (Eds.), . ASM Press, Washington, DC, pp. 403-422.

Reznikoff, W.S., 2008. Transposon Tn*5*. Annu. Rev. Genet. 42, 269-286.

Reznikoff, W.S., Jilk, R., Krebs, M.P., Makris, J.C., Nordmann, P.L., Weinreich, M., Wiegand, T., 1993. Tn*5* lacZ translation fusion element: isolation and analysis of transposition mutants. Methods Enzymol. 217, 312-322.

Rezsohözy, R., Hallet, B., Delcour, J., Mahillon, J., 1993. The IS*4* family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. 9, 1283-1295.

Rice, P., Craigie, R., Davies, D.R., 1996. Retroviral integrases and their cousins. Curr. Opin. Struct. Biol. 6, 76-83.

Rice, P., Mizuuchi, K., 1995. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. Cell 82, 209-220.

Rice, P.A., Baker, T.A., 2001. Comparative architecture of transposase and integrase complexes. Nat. Struct. Biol. 8, 302-307.

Richardson, J.M., Colloms, S.D., Finnegan, D.J., Walkinshaw, M.D., 2009. Molecular architecture of the *Mos1* paired-end complex: the structural basis of DNA transposition in a eukaryote. Cell 138, 1096-1108.

Rothenberg, S.M., Olsen, M.N., Laurent, L.C., Crowley, R.A., Brown, P.O., 2001. Comprehensive mutational analysis of the Moloney murine leukemia virus envelope protein. J. Virol. 75, 11851-11862.

Sakai, J., Kleckner, N., 1997. The Tn*10* synaptic complex can capture a target DNA only after transposon excision. Cell 89, 205-214.

Savilahti, H., Mizuuchi, K., 1996. Mu transpositional recombination: donor DNA cleavage and strand transfer *in trans* by the Mu transposase. Cell 85, 271-280.

Savilahti, H., Rice, P.A., Mizuuchi, K., 1995. The phage Mu transpososome core: DNA requirements for assembly and function. EMBO J. 14, 4893-4903.

Schumacher, S., Clubb, R.T., Cai, M., Mizuuchi, K., Clore, G.M., Gronenborn, A.M., 1997. Solution structure of the Mu end DNA-binding Iβ subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. EMBO J. 16, 7532-7541.

Schweidenback, C.T., Baker, T.A., 2008. Dissecting the roles of MuB in Mu transposition: ATP regulation of DNA binding is not essential for target delivery. Proc. Natl. Acad. Sci. U. S. A. 105, 12101-12107.

Shapiro, J.A., 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. U. S. A. 76, 1933-1937.

Shapiro, J.A., 2010. Mobile DNA and evolution in the 21st century. Mob DNA 1, 4-8753-1-4.

Shivange, A.V., Marienhagen, J., Mundhada, H., Schenk, A., Schwaneberg, U., 2009. Advances in generating functional diversity for directed protein evolution. Curr. Opin. Chem. Biol. 13, 19-25.

Siegele, D.A., Hu, J.C., 1997. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. Proc. Natl. Acad. Sci. U. S. A. 94, 8168-8172.

Siguier, P., Filee, J., Chandler, M., 2006a. Insertion sequences in prokaryotic genomes. Curr. Opin. Microbiol. 9, 526-531.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., Chandler, M., 2006b. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34, D32-6.

Singer, B., Kusmierek, J.T., 1982. Chemical mutagenesis. Annu. Rev. Biochem. 51, 655-693.

Singh, I.R., Crowley, R.A., Brown, P.O., 1997. High-resolution functional mapping of a cloned gene by genetic footprinting. Proc. Natl. Acad. Sci. U. S. A. 94, 1304-1309.

Snyder, L., Champness, W., 2007. Molecular Genetics of Bacteria. ASM Press, Washington, D. C., USA.

Steiniger-White, M., Rayment, I., Reznikoff, W.S., 2004. Structure/function insights into Tn5 transposition. Curr. Opin. Struct. Biol. 14, 50-57.

Stellwagen, A.E., Craig, N.L., 1997a. Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn*7*. Genetics 145, 573-585.

Stellwagen, A.E., Craig, N.L., 1997b. Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn*7* . Genetics 145, 573-585.

Stemmer, W.P., 1994a. Rapid evolution of a protein *in vitro* by DNA shuffling . Nature 370, 389-391.

Stemmer, W.P., 1994b. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. U. S. A. 91, 10747-10751.

Stratagene, 2004. Overcome mutational bias. Strategies 17, 20-21.

Surette, M.G., Buch, S.J., Chaconas, G., 1987. Transpososomes: stable protein-DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. Cell 49, 253-262.

Surette, M.G., Chaconas, G., 1991. Stimulation of the Mu DNA strand cleavage and intramolecular strand transfer reactions by the Mu B protein is independent of stable binding of the Mu B protein to DNA. J. Biol. Chem. 266, 17306-17313.

Surette, M.G., Chaconas, G., 1992. The Mu transpositional enhancer can function *in trans*: requirement of the enhancer for synapsis but not strand cleavage. Cell 68, 1101-1108.

Surette, M.G., Lavoie, B.D., Chaconas, G., 1989. Action at a distance in Mu DNA transposition: an enhancer-like element is the site of action of supercoiling relief activity by integration host factor (IHF). EMBO J. 8, 3483-3489.

Swingle, B., O'Carroll, M., Haniford, D., Derbyshire, K.M., 2004. The effect of host-encoded nucleoid proteins on transposition: H-NS influences targeting of both IS*903* and Tn*10*. Mol. Microbiol. 52, 1055-1067.

Taira, S., Tuimala, J., Roine, E., Nurmiaho-Lassila, E.L., Savilahti, H., Romantschuk, M., 1999. Mutational analysis of the Pseudomonas syringae pv. tomato hrpA gene encoding Hrp pilus subunit. Mol. Microbiol. 34, 737-744.

Tang, Y., Cotterill, S., Lichtenstein, C.P., 1995. Genetic analysis of the terminal 8-bp inverted repeats of transposon Tn7. Gene 162, 41-46.

Tavakoli, N.P., Derbyshire, K.M., 1999. IS*903* transposase mutants that suppress defective inverted repeats. Mol. Microbiol. 31, 1183-1195.

Tavakoli, N.P., Derbyshire, K.M., 2001. Tipping the balance between replicative and simple transposition. EMBO J. 20, 2923-2930.

Taylor, A.L., 1963. Bacteriophage-Induced Mutation in *Escherichia Coli*. Proc. Natl. Acad. Sci. U. S. A. 50, 1043-1051.

Tu Quoc, P.H., Genevaux, P., Pajunen, M., Savilahti, H., Georgopoulos, C., Schrenzel, J., Kelley, W.L., 2007. Isolation and characterization of biofilm formation-defective mutants of *Staphylococcus aureus*. Infect. Immun. 75, 1079-1088.

Turakainen, H., Saarimaki-Vire, J., Sinjushina, N., Partanen, J., Savilahti, H., 2009. Transposition-based method for the rapid generation of gene-targeting vectors to produce Cre/Flp-modifiable conditional knock-out mice. PLoS One 4, e4341.

Turlan, C., Chandler, M., 2000. Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. Trends Microbiol. 8, 268-274.

Twiss, E., Coros, A.M., Tavakoli, N.P., Derbyshire, K.M., 2005. Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress. Mol. Microbiol. 57, 1593-1607.

van Luenen, H.G., Plasterk, R.H., 1994. Target site choice of the related transposable elements Tc1 and Tc3 of *Caenorhabditis elegans*. Nucleic Acids Res. 22, 262-269.

van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H., Sixma, T.K., 1997. Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C.elegans* in complex with transposon DNA. EMBO J. 16, 6044-6054.

Vilen, H., Aalto, J.M., Kassinen, A., Paulin, L., Savilahti, H., 2003. A direct transposon insertion tool for modification and functional analysis of viral genomes. J. Virol. 77, 123-134.

Vilen, H., Eerikainen, S., Tornberg, J., Airaksinen, M.S., Savilahti, H., 2001. Construction of gene-targeting vectors: a rapid Mu in vitro DNA transposition-based strategy generating null, potentially hypomorphic, and conditional alleles. Transgenic Res. 10, 69-80.

Volff, J.N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28, 913-922.

Volles, M.J., Lansbury, P.T.,Jr, 2005. A computer program for the estimation of protein and nucleic acid sequence diversity in random point mutagenesis libraries. Nucleic Acids Res. 33, 3667-3677.

Waddell, C.S., Craig, N.L., 1988. Tn*7* transposition: two transposition pathways directed by five Tn*7*-encoded genes. Genes Dev. 2, 137-149.

Wang, Z., Harshey, R.M., 1994. Crucial role for DNA supercoiling in Mu transposition: a kinetic study. Proc. Natl. Acad. Sci. U. S. A. 91, 699-703.

Wang, Z., Namgoong, S.Y., Zhang, X., Harshey, R.M., 1996. Kinetic and structural probing of the precleavage synaptic complex (type 0) formed during phage Mu transposition. Action of metal ions and reagents specific to single-stranded DNA. J. Biol. Chem. 271, 9619-9626.

Watkins, S., van Pouderoyen, G., Sixma, T.K., 2004. Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. Nucleic Acids Res. 32, 4306-4312.

Watson, M.A., Chaconas, G., 1996. Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. Cell 85, 435-445.

Weber, M., Chernov, K., Turakainen, H., Wohlfahrt, G., Pajunen, M., Savilahti, H., Jantti, J., 2010. Mso1p Regulates Membrane Fusion through Interactions with the Putative N-Peptide-binding Area in Sec1p Domain 1. Mol. Biol. Cell 21, 1362-1374.

Weinreich, M.D., Gasch, A., Reznikoff, W.S., 1994. Evidence that the *cis* preference of the Tn*5* transposase is caused by nonproductive multimerization. Genes Dev. 8, 2363-2374.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973-982.

Wiegand, T.W., Reznikoff, W.S., 1992. Characterization of two hypertransposing Tn*5* mutants. J. Bacteriol. 174, 1229-1239.

Williams, T.L., Jackson, E.L., Carritte, A., Baker, T.A., 1999. Organization and dynamics of the Mu transpososome: recombination by communication between two active sites. Genes Dev. 13, 2725-2737.

Wong, T.S., Roccatano, D., Loakes, D., Tee, K.L., Schenk, A., Hauer, B., Schwaneberg, U., 2008. Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR. Biotechnol. J. 3, 74-82.

Wong, T.S., Roccatano, D., Schwaneberg, U., 2007a. Are transversion mutations better? A Mutagenesis Assistant Program analysis on P450 BM-3 heme domain. Biotechnol. J. 2, 133-142.

Wong, T.S., Roccatano, D., Schwaneberg, U., 2007b. Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries. Environ. Microbiol. 9, 2645-2659.

Wong, T.S., Roccatano, D., Zacharias, M., Schwaneberg, U., 2006. A statistical analysis of random mutagenesis methods used for directed protein evolution. J. Mol. Biol. 355, 858-871.

Wong, T.S., Tee, K.L., Hauer, B., Schwaneberg, U., 2004. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. Nucleic Acids Res. 32, e26.

Wong, T.S., Zhurina, D., Schwaneberg, U., 2006. The diversity challenge in directed protein evolution. Comb. Chem. High Throughput Screen. 9, 271-288.

Wu, Z., Chaconas, G., 1992. Flanking host sequences can exert an inhibitory effect on the cleavage step of the in vitro mu DNA strand transfer reaction. J. Biol. Chem. 267, 9552-9558.

Wu, Z., Chaconas, G., 1995. A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage. EMBO J. 14, 3835-3843.

Wu, Z., Xuanyuan, Z., Li, R., Jiang, D., Li, C., Xu, H., Bai, Y., Zhang, X., Turakainen, H., Saris, P.E., Savilahti, H., Qiao, M., 2009. Mu transposition complex mutagenesis in *Lactococcus lactis*--identification of genes affecting nisin production. J. Appl. Microbiol. 106, 41-48.

Yanagihara, K., Mizuuchi, K., 2002. Mismatch-targeted transposition of Mu: a new strategy to map genetic polymorphism. Proc. Natl. Acad. Sci. U. S. A. 99, 11317-11321.

Yanagihara, K., Mizuuchi, K., 2003. Progressive structural transitions within Mu transpositional complexes. Mol. Cell 11, 215-224.

Yang, J.Y., Jayaram, M., Harshey, R.M., 1995. Enhancer-independent variants of phage Mu transposase: enhancer-specific stimulation of catalytic activity by a partner transposase. Genes Dev. 9, 2545-2555.

Yang, J.Y., Jayaram, M., Harshey, R.M., 1996. Positional information within the Mu transposase tetramer: catalytic contributions of individual monomers. Cell 85, 447-455.

Yang, J.Y., Kim, K., Jayaram, M., Harshey, R.M., 1995. A domain sharing model for active site assembly within the Mu A tetramer during transposition: the enhancer may specify domain contributions. EMBO J. 14, 2374-2384.

Yang, W., Hendrickson, W.A., Crouch, R.J., Satow, Y., 1990. Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. Science 249, 1398-1405.

Yang, W., Lee, J.Y., Nowotny, M., 2006. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol. Cell 22, 5-13.

Yin, Z., Jayaram, M., Pathania, S., Harshey, R.M., 2005. The Mu transposase interwraps distant DNA sites within a functional transpososome in the absence of DNA supercoiling. J. Biol. Chem. 280, 6149-6156.

Yin, Z., Suzuki, A., Lou, Z., Jayaram, M., Harshey, R.M., 2007. Interactions of phage Mu enhancer and termini that specify the assembly of a topologically unique interwrapped transpososome. J. Mol. Biol. 372, 382-396.

Yuan, J.F., Beniac, D.R., Chaconas, G., Ottensmeyer, F.P., 2005a. 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. Genes Dev. 19, 840-852.

Yuan, L., Kurek, I., English, J., Keenan, R., 2005b. Laboratory-directed protein evolution. Microbiol. Mol. Biol. Rev. 69, 373-392.

Yuan, Y.W., Wessler, S.R., 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. Proc. Natl. Acad. Sci. U. S. A. 108, 7884-7889.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., Craig, N.L., 2011. A hyperactive *piggyBac* transposase for mammalian applications. Proc. Natl. Acad. Sci. U. S. A. 108, 1531-1536.

Zaccolo, M., Williams, D.M., Brown, D.M., Gherardi, E., 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J. Mol. Biol. 255, 589-603.

Zamudio, N., Bourc'his, D., 2010. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap?. Heredity (Edinb) 105, 92-104.

Zayed, H., Izsvák, Z., Walisko, O., Ivics, Z., 2004. Development of hyperactive *Sleeping Beauty* transposon vectors by mutational analysis. Mol. Ther. 9, 292-304.

Zhang, C., Kitsberg, D., Chy, H., Zhou, Q., Morrison, J.R., 2005. Transposon-mediated generation of targeting vectors for the production of gene knockouts. Nucleic Acids Res. 33, e24.

Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F., Craig, N.L., 2004. Transposition of *hAT* elements links transposable elements and V(D)J recombination. Nature 432, 995-1001.

Zou, A.H., Leung, P.C., Harshey, R.M., 1991. Transposase contacts with Mu DNA ends. J. Biol. Chem. 266, 20476-20482.